

Building a Predictive Model of Galaxy Formation - I: Phenomenological Model Constrained to the $z = 0$ Stellar Mass Function

Andrew J. Benson

Carnegie Observatories, 813 Santa Barbara Street, Pasadena, CA 91101, USA.

23 May 2014

ABSTRACT

We constrain a highly simplified semi-analytic model of galaxy formation using the $z \approx 0$ stellar mass function of galaxies. Particular attention is paid to assessing the role of random and systematic errors in the determination of stellar masses, to systematic uncertainties in the model, and to correlations between bins in the measured and modeled stellar mass functions, in order to construct a realistic likelihood function. We derive constraints on model parameters and explore which aspects of the observational data constrain particular parameter combinations. We find that our model, once constrained, provides a remarkable match to the measured evolution of the stellar mass function to $z = 1$, although fails dramatically to match the local galaxy HI mass function. Several “nuisance parameters” contribute significantly to uncertainties in model predictions. In particular, systematic errors in stellar mass estimate are the dominant source of uncertainty in model predictions at $z \approx 1$, with additional, non-negligible contributions arising from systematic uncertainties in halo mass functions and the residual uncertainties in cosmological parameters. Ignoring any of these sources of uncertainties could lead to viable models being erroneously ruled out. Additionally, we demonstrate that ignoring the significant covariance between bins the observed stellar mass function leads to significant biases in the constraints derived on model parameters. Careful treatment of systematic and random errors in the constraining data, and in the model being constrained, are crucial if this methodology is to be used to test hypotheses relating to the physics of galaxy formation.

Key words: galaxies: evolution, galaxies: formation

1 INTRODUCTION

Galaxy formation is a complex, nonlinear process, driven by the interplay of many different physical mechanisms (Benson 2010). Fortunately, there exists a wealth of observational data from which we can attempt to infer the physics of galaxy formation. Given the complexity of the problem, this inference necessarily proceeds via the use of models, which we use to estimate the statistical properties of the galaxy population given some set of assumptions about the underlying theory of galaxy formation. The goal of this process is to proceed from information (observational data), through knowledge (empirical descriptions of the galaxy population which summarize observational facts), to understanding (insight into the underlying physics).

Recently, there has been significant interest in the process of quantitatively constraining the parameters of semi-analytic models (SAMs) using Bayesian Markov Chain Monte Carlo (MCMC) (Henriques et al. 2009; Henriques &

Thomas 2010; Lu et al. 2011, 2012; Mutch et al. 2013; Lu et al. 2013), model emulator (Bower et al. 2010; Gómez et al. 2012, 2013) methods, and particle swarm algorithms¹ (Ruiz et al. 2013). This approach is crucial to facilitate the goal of model inference—that is, inferring the likelihood of a model given the data—and for making the reliable, quantitative predictions that are necessary in order to test models. To fully exploit the power of this approach it is crucial to address the following issues:

- (i) construction of realistic likelihood functions;

¹ Note that, unlike MCMC and model emulator techniques, the particle swarm method does not directly compute the full posterior probability distribution of the model parameters. Instead the posterior probability distribution must be approximate as a multivariate Gaussian by considering the behaviour of points in the vicinity of the maximum likelihood solution.

- (ii) accounting for random and systematic errors in observations *and* models;
- (iii) inferring the minimalistic viable model.

The first issue requires careful assessment of the covariance in both observational and model determinations of the statistic (e.g. galaxy mass function) in question. The second issue requires a careful consideration of all sources of error in the observations, and model discrepancies (i.e. differences between the model and the modeled reality arising from approximations inherent in the modeling). Finally, the third point requires that we begin by studying simplistic models and gradually add complexity as guided by the data.

1.1 Goals

This is the first in a series of papers in which we will develop a well-constrained SAM of galaxy formation based on the GALACTICUS toolkit². Specifically, we wish to constrain a model of galaxy formation in the standard Λ CDM cosmological model. We could, alternatively, choose to relax this assumption and explore non-flat cosmologies, with dark energy with equation of state $w \neq -1$, and with a variety of dark matter particle phenomenology (warm dark matter, self-interacting dark matter, etc.). We choose *not* to do this for two reasons. First, such an exploration is currently computationally impractical³. The second reason is that our understanding of galaxy formation remains sufficiently poor that we do not expect it to give strong insights into the nature of these extensions to the standard cosmological model. As such, we believe that these extensions should only be explored in studies such as this work if they are strongly motivated by other experiments. Currently, this is not the case (Frenk & White 2012; Hinshaw et al. 2013).

Our goal is to construct a model of galaxy formation by starting with the simplest possible model and adding physics, guided at each step by constraints from the best available observations of key measures of the galaxy population. We intend to put a very significant emphasis on carefully assessing uncertainties, both random and systematic, in both observations and model, so as avoid being misled by over-constraining the model, or constraining to biased data.

1.2 Motivation for Simple Models

The GALACTICUS model can incorporate detailed modeling of galaxy formation physics (Benson 2012). However, for both practical and methodological reasons, we choose in this work to run GALACTICUS with a highly simplified model. Specifically, we run a model that contains minimal physics beyond that of dark matter structure formation and subhalo merging, instead relying entirely on empirical prescriptions.

Practically, such a simplified model runs much faster than a model with the complete complement of physics, making the MCMC analysis tractable.

² <https://sites.google.com/site/galacticusmodel/>

³ While GALACTICUS is able to compute solutions for all of these variants on the standard cosmological model, the additional parameters and computational load introduced make it impractical for MCMC exploration at present.

From the standpoint of methodology, we wish to build up a complete description of galaxy formation by gradually piecing together the components of a model, guided at each step by the best available observational data. Obviously, this must be a physical model of galaxy formation. However, our initial goal is simply to construct a prescription close enough to basic data upon which we can then begin to improve. Additionally, it is interesting to understand what the basic scaling properties of different processes must be in order to reproduce data. From a model inference perspective, we also wish to begin with a model that we *expect* to be able to rule out. When the model is ruled out, we will switch one or more components to a more physical model, or add additional physics, until a viable model is recovered.

Our simple model is therefore constructed to encapsulate the following established general principles of galaxy formation theory (Benson 2010):

- gas accretes into dark matter halos from the intergalactic medium (IGM);
- that gas, initially diffusely distributed throughout the halo, must cool⁴ and condense to galactic densities;
- stars must form from this condensed gas;
- some material must be driven out of galaxies to limit star formation in lower mass halos;
- above some halo mass, some additional process must act to inhibit the condensation of gas (Benson et al. 2003).

Based on the success of similar prescriptions (e.g. Neistein & Weinmann 2010), we expect that such a model, carefully constructed, should be able to match the constraint that we will apply in this work.

1.3 Choice of Constraint

We have chosen as our initial, sole constraint, the stellar mass function of galaxies measured in the local Universe, specifically the measurement by Li & White (2009). While there are more recent measures of the stellar mass function (which appeared after this study was begun, and some of which are explored in §4), the precise choice of dataset should not matter in the following sense. If all sources of random and systematic error affecting the data are correctly accounted for then constraints on the model parameters should be unbiased, and therefore consistent with constraints that would be derived from more recent measurements. Of course, more recent measurements might have smaller random and systematic errors, and thereby provide stronger constraints—this will be invaluable to future improvements in our modeling, but our primary goal here is to derive robust, unbiased constraints on model parameters.

The stellar mass function represents one of the most fundamental properties of the galaxy population that a viable theory of galaxy formation should be able to reproduce. There are, of course, other fundamental properties that that theory should also reproduce or predict (e.g. luminosity functions, gas content, star formation rates, the dependence of galaxy clustering on luminosity etc.). Our goal

⁴ Or it may already be cool (Birnboim & Dekel 2003; Kereš et al. 2005)—in any case it must collect in a high density region at the center of the halo where it can form a galaxy.

here is to select just a single, fundamental measure of the galaxy population. In future works in this series we will then explore the effects of adding additional constraints. Furthermore, many of these (e.g. luminosity functions and star formation rates) are closely tied to the stellar mass function.

The remainder of this paper is arranged as follows. In §2 we describe the galaxy formation model that will be employed in this work. §3 describes how we construct the model likelihood function, including random and systematic errors in model and data. In §4 we present results on the parameter constraints derived, and explore extrapolations of the model to other datasets. These results are discussed in §5, and we present our overall conclusions in §6. We include several appendices addressing various technical aspects of our modeling. Appendix A describes how we build a halo occupation distribution (HOD) model of the observed stellar mass function (see also §3.1), Appendix B describes an algorithm for selecting the optimal set of halo masses to simulate in a SAM when constraining to a given data set, Appendix C explores the numerical robustness and convergence of our SAM code, Appendix D describes our model for the covariance matrix of the model galaxy stellar mass function, and, finally, Appendix E explores the accuracy of the algorithm we employ for merger tree construction.

2 MODEL

As stated in §1, our aim is to begin with a highly simplified, phenomenological model of galaxy formation which we will refine (and into which we will introduce more physics) as aspects of the model are ruled out by consideration of a variety of constraints. Therefore, we begin with a model which aims to compute only the distribution of galaxy masses (both stellar and gaseous) using simple scaling relations applied to dark matter halo merger trees.

We note that the model described below is not the simplest possible model. One could imagine a model which neglected the role of cooling and simply assumed a star formation rate proportional to the growth rate of halos. This would result in a stellar mass function that was a simple mapping of the halo(+subhalo) mass function to lower masses. It is well known that such a model is highly incompatible with observations (Benson et al. 2003). Furthermore, we could consider a model with no feedback. Such a model might plausibly fit the $z = 0$ stellar mass function given a sufficiently flexible prescription for the star formation rate⁵.

Instead, we adopt a *sufficiently simple* model, by which we mean one in which the various prescriptions are simple scaling relations, and the interconnections between the

mass reservoirs are simple to understand. We include feedback specifically so that we can explore whether feedback is actually required by the constraints for such a sufficiently simple phenomenological model. This model can be run using v0.9.3 of GALACTICUS (revision 2284).

Our sufficiently simple model tracks the flow of mass between three reservoirs:

- *hot halo gas* defined as diffuse gas occupying the majority of the volume of the dark matter halo (a.k.a. circumgalactic medium (CGM), intracluster medium (ICM));
- *ISM gas* defined as gas in the interstellar medium (ISM) of a galaxy;
- *stars* defined as mass which is locked up into long-lived (main sequence lifetimes $\gtrsim 10$ Gyr) stars.

The mass in each component for a given halo is denoted by M_{hot} , M_{ISM} , and M_{\star} respectively.

Our model is fully specified by three classes of rule:

- (i) dark matter merger trees;
- (ii) a set of ordinary differential equations (ODEs) describing the evolution of the baryonic reservoirs in isolated galaxies;
- (iii) rules for the effects of halo and galaxy mergers on the baryonic reservoirs.

We proceed to detail these three classes below.

2.1 Merger Trees

Merger trees are generated using the algorithm described by Parkinson et al. (2008), which reproduces the progenitor mass functions of halos in the Millennium Simulation. Jiang & van den Bosch (2013) study several algorithms for constructing merger trees and compare their results with trees extracted from N-body simulations. They find that the Parkinson et al. (2008) algorithm is the only algorithm which gives good agreement with the N-body merger trees for all statistics that they consider. Trees are generated backward in time starting from $z = 0$, and are followed until each branch falls below a specified mass resolution, M_{res} . We test for convergence in our results with respect to the numerical parameters controlling the construction of these trees (see Appendix C). Throughout, we use the fitting formulae of Kitayama & Suto (1996) to compute the virial density contrast for dark matter halos, Δ_{h} , and the critical overdensity for collapse, δ_{c} .

In the interests of computational speed, we allow the mass resolution to be a function of the final, $z = 0$, mass of the tree, M_0 . Specifically,

$$M_{\text{res}} = \max(f_0 M_0, M_{\text{res,min}}), \quad (1)$$

where $f_0 = 0.1$ and $M_{\text{res,min}} = 10^8 M_{\odot}$. This approximation will introduce some error into the model calculation. We measure and account for this error as a model discrepancy term (see §3.5.4).

We must also specify the set of $z = 0$ halo masses for which trees are to be built. These masses can be distributed arbitrarily in the range of required masses, providing the contribution of their galaxies to the stellar mass function is appropriately weighted. Given a set of $z = 0$ halo masses,

⁵ For example, if the reduced specific star formation rate is arranged to be twice the halo growth rate, $\dot{M}_{\star}/M_{\star} = 2\dot{M}_{\text{h}}/M_{\text{h}}$, then $M_{\star} \propto M_{\text{h}}^2$ which will result in a flattening of the low-mass stellar mass function relative to the low-mass halo mass function, without any explicit inclusion of feedback. Indeed, the model of Lilly et al. (2013) succeeds in matching the stellar mass function for precisely this reason. It should be noted however that this is not a physical model of star formation. As such, this phenomenological prescription for the star formation rate may well be encapsulating the net effect of the physics of star formation plus feedback.

M_i , the weight to assign galaxies in tree i is simply

$$w_i = \int_{\sqrt{M_{i+1}/M_i}}^{\sqrt{M_i/M_{i-1}}} \minmax \left[\phi_{\min}, \phi_{\max}, \frac{dn}{d \ln M}(M) \right] M dM, \quad (2)$$

where $dn/d \ln M$ is the halo mass function, for which we use the fitting function of Tinker et al. (2008) with the appropriate value of Δ_h , and

$$\minmax(a, b, x) = \begin{cases} a & \text{if } x < a \\ x & \text{if } a \leq x \leq b \\ b & \text{if } x > b, \end{cases} \quad (3)$$

with $\phi_{\min} = 10^{-7} \text{ Mpc}^{-3}$ and $\phi_{\max} = 10^{-3} \text{ Mpc}^{-3}$. The choice of distribution of $z = 0$ halo masses will affect both the precision of results obtained (since we construct the stellar mass function by summing over realizations of merger trees, the finite number of trees generated will lead to statistical fluctuations in the results), and the time taken for each model evaluation. In Appendix B we explore three different distributions of $z = 0$ halo masses and find that simply constructing a volume-limited sample (i.e. drawing masses at random from the $z = 0$ halo mass function) is close to optimal.

2.1.1 Cosmological Dependence

The statistics of the merger tree structure will depend on the cosmological parameters (via the critical overdensity for collapse, δ_c), which we will allow to vary in our analysis. Additionally, given the algorithm described above, the $z = 0$ masses of the trees will depend both on the cosmological parameters and on systematic uncertainties in the halo mass function (see §3.3). Unfortunately, this dependence can introduce discontinuities in the likelihood surface. Since tree construction involves binary branching a small change in a cosmological parameter can lead to large changes in the structure of one or more trees. (Similarly, changing the $z = 0$ mass by a small amount can lead to a large change in the structure of a tree.)

In principle this is not a problem, providing that the number of trees simulated is very large—the more trees simulated the smaller the discontinuities in the likelihood surface will be. In practice, however, the number of trees we can run (given computational constraints) is not that large, and the discontinuities in the likelihood surface are significant, making it difficult for an MCMC algorithm to achieve convergence.

Therefore, we adopt the following approach. A single set of merger trees is generated using the maximum likelihood cosmological parameters from (Hinshaw et al. 2013, see §3.7.2 for a full discussion of how we utilize these cosmological parameter constraints) and no systematic offsets in the halo mass function. These trees are stored to file, and are re-used by each model evaluation⁶. The weight assigned to each tree (see eqn. 2) is recomputed for each model evaluation using the cosmological parameters of that model.

⁶ Specifically, the masses and redshifts of each node in each tree is reused in model evaluations. This will lead to some difference in the time at which nodes are defined since the cosmological time-redshift relation will vary between model evaluations.

This approach reduces variance between model evaluations—a point that we will address in §3.5.5.

We note that, in future MCMC applications where a very large number of trees can be used, there is no reason to use a fixed set of trees, even if the model is evaluated twice with exactly the same cosmological parameters. While this does make the likelihood surface stochastic this would not affect the assumption of detailed balance in the MCMC algorithm, since the probability for the merger-tree process is Lebesgue measurable⁷, but would broaden the posterior probability distribution for the model parameters⁸. This approach is desirable in principle, as it avoids any bias in the model posterior which may be introduced through the use of a single set of merger trees (i.e. any finite set of merger trees will be somewhat biased relative to the full ensemble).

We have tested this approach, building a unique set of merger trees each time the model is evaluated, and find that at present it significantly extends the convergence time for the MCMC algorithm, making it currently impractical.

2.2 Differential Evolution

Galaxies in our model typically evolve differentially, except when this evolution is punctuated by merger events (see §2.3). The set of ODEs to be solved along each branch of every merger tree is described below.

2.2.1 Accretion From The IGM

Gas is assumed to accrete into halos from the IGM at a rate proportional to the growth rate of the total halo mass, with the constant of proportionality being the universal baryon fraction. The exception to this rule occurs for halos deemed to be unable to accrete due to the the IGM being heated to high temperatures as a result of the reionization of the Universe. We adopt a simple step-function model for this suppression of accretion due to reionization which has been shown to capture the essential features of more realistic models (Font et al. 2011). Specifically:

$$\dot{M}_{\text{IGM}} = \begin{cases} (\Omega_b/\Omega_M) \dot{M}_{\text{h,smooth}} & \text{if } V_h > V_{\text{reion}} \text{ or } z > z_{\text{reion}}, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where Ω_b and Ω_M are the density parameters for baryons and total matter respectively, $\dot{M}_{\text{h,smooth}}$ is the total rate⁹ of smooth accretion onto the halo¹⁰, V_h is the halo virial velocity, z is redshift, and V_{reion} and z_{reion} are parameters of the model.

⁷ We thank Martin Weinberg for clarifying this point.

⁸ This broadening is instead accounted for by including our estimate of the model uncertainty in the likelihood function as described in Appendix D.

⁹ As determined from our merger trees.

¹⁰ Hot gas will also be obtained from merging with resolved halos—see §2.3.1.

2.2.2 Cooling

The cooling of gas¹¹ from the hot halo to the ISM of a galaxy is modeled with a simple redshift-dependent timescale¹², plus an exponential break¹³:

$$\dot{M}_{\text{cool}} = \frac{M_{\text{hot}}}{\tau_{\text{cool}}} (1+z)^{-\alpha_{\text{cool}}} \times \left(1 + \exp \left[\frac{\log_{10}(M_{\text{halo}}/\mathcal{M}_{\text{cool}})}{\Delta \log_{10} \mathcal{M}_{\text{cool}}} \right] \right)^{-\beta_{\text{cool}}}, \quad (5)$$

where τ_{cool} , α_{cool} , β_{cool} , $\mathcal{M}_{\text{cool}}$, and $\Delta \log_{10} \mathcal{M}_{\text{cool}}$ are parameters of the model.

2.2.3 Star Formation

The rate of star formation is assumed to be proportional to the mass of the ISM, and inversely proportional to a timescale that scales with both redshift and halo virial velocity as follows:

$$\dot{M}_{\text{sf}} = \min \left(\frac{M_{\text{ISM}}}{\tau_{\star}} (1+z)^{-\alpha_{\star}} \left[\frac{V_{\text{h}}}{200 \text{ km/s}} \right]^{-\beta_{\star}}, \frac{M_{\text{ISM}}}{\tau_{\star, \text{min}}} \right), \quad (6)$$

where τ_{\star} , α_{\star} , β_{\star} , and $\tau_{\star, \text{min}}$ are parameters of the model (the final parameter being introduced to prevent timescales becoming arbitrarily small). Note that we do not explicitly consider recycling of material from stars. The star formation rate should therefore be viewed as the *reduced* star formation rate (i.e. the rate of formation of long-lived stars).

2.2.4 Feedback

Feedback (i.e. outflow of material from the ISM to the hot halo) is assumed to occur at a rate proportional to the star formation rate, with a constant of proportionality which

¹¹ More specifically, since we do not include any physics here, this should be viewed as the *transfer* of gas from the hot halo to the ISM. Whether or not this occurs via cooling is irrelevant.

¹² Here, and elsewhere, we introduce an explicit redshift-dependence into our model. It could be argued that this is unphysical since a galaxy should have no direct knowledge of the redshift (expect, perhaps, via a redshift-dependent background radiation field for example). However, we justify the inclusion of such a explicit dependence by arguing that the evolution of a galaxy may depend directly on its density, which is expected to scale with the density of the Universe—a redshift-dependent quantity. Our present, sufficiently simple model does not directly trace galaxy densities, so we believe inclusion of an explicit redshift dependence as a proxy is justified. Of course, our goal is to proceed toward more physical models in which any such density dependence is modeled directly.

¹³ The exponential break is introduced since it is well-established that without this type of suppression it is impossible to reproduce the exponential decline in the abundances of high-mass galaxies (Benson et al. 2003). We expect, therefore, that our MCMC analysis will show that the parameter $\mathcal{M}_{\text{cool}}$ cannot be arbitrarily large.

scales with both redshift and halo virial velocity:

$$\dot{M}_{\text{wind}} = \min \left(f_{\text{wind}} \dot{M}_{\text{sf}} (1+z)^{\alpha_{\text{wind}}} \left[\frac{V_{\text{h}}}{200 \text{ km/s}} \right]^{\beta_{\text{wind}}}, \frac{M_{\text{ISM}}}{\tau_{\text{wind}, \text{min}}} \right), \quad (7)$$

where f_{wind} , α_{wind} , β_{wind} , and $\tau_{\text{wind}, \text{min}}$ are parameters of the model (the final parameter being introduced to prevent arbitrarily high outflow rates from occurring).

2.2.5 Network

Finally, the rates described above are combined into the following network describing the net rates of change of the masses in our three reservoirs:

$$\dot{M}_{\text{hot}} = \dot{M}_{\text{IGM}} - \dot{M}_{\text{cool}} + \dot{M}_{\text{wind}}, \quad (8)$$

$$\dot{M}_{\text{ISM}} = \dot{M}_{\text{cool}} - \dot{M}_{\text{wind}} - \dot{M}_{\text{sf}}, \quad (9)$$

$$\dot{M}_{\star} = \dot{M}_{\text{sf}}. \quad (10)$$

Note that the only net source of mass is accretion from the IGM. This network of ODEs is integrated along each branch of each merger tree.

2.2.6 Initial Conditions

The initial conditions for differential evolution of our model are given by specifying the baryonic content of dark matter halos corresponding to the tip (i.e. progenitorless halo) of each branch in the merger tree. We assume that each branch-tip halo contains zero mass in ISM and stellar mass, and a mass of hot gas equal to

$$M_{\text{hot}} = \begin{cases} (\Omega_{\text{b}}/\Omega_{\text{M}}) M_{\text{h}} & \text{if } V_{\text{h}} > V_{\text{reion}} \text{ or } z > z_{\text{reion}}, \\ 0 & \text{otherwise,} \end{cases} \quad (11)$$

where M_{h} is the total mass of the halo.

2.3 Punctuated Evolution

In addition to the differential evolution of galaxies along each branch of the merger tree, we include instantaneous, punctuated evolution in response to mergers of both halos and galaxies, as described below.

2.3.1 Halo Merging

When two halos merge (i.e. when a smaller halo first becomes a subhalo in a larger host halo) the hot halo gas from the new subhalo is assumed to be instantaneously removed¹⁴ and is added to the hot halo gas of the host halo:

$$\begin{aligned} M_{\text{hot}, \text{host}} &\rightarrow M_{\text{hot}, \text{host}} + M_{\text{hot}, \text{subhalo}}; \\ M_{\text{hot}, \text{subhalo}} &\rightarrow 0. \end{aligned} \quad (12)$$

Additionally, the subhalo is assigned a time-of-merging, t_{merge} , at which it will undergo a galaxy merger with the central galaxy of its host halo. The time-of-merging is computed

¹⁴ Presumably by ram pressure or other environmental forces, although the details are purposely not specified in this sufficiently simple model.

using $t_{\text{merge}} = t + t_{\text{df}}$ where t is the time at which the halo becomes a subhalo, and t_{df} is the dynamical friction timescale computed using the fitting formula of (Jiang et al. 2008, their eqn. 5), accounting for both mass and orbit dependencies. The orbital parameters of each new subhalo are assumed to be precisely equal to the mode of the distribution found by Benson (2005), specifically $(V_r, V_\phi) = (0.90, 0.75)V_{h,\text{host}}$, where $V_{h,\text{host}}$ is the virial velocity of the host halo.

2.3.2 Galaxy Merging

At time t_{merge} , the subhalo is removed from the calculation. The masses of its baryonic reservoirs (ISM and stars only, as its hot halo gas reservoir has already been removed) are simply added to those of the central galaxy of the subhalo's host halo:

$$\begin{aligned} M_{\text{ISM},\text{host}} &\rightarrow M_{\text{ISM},\text{host}} + M_{\text{ISM},\text{subhalo}}, \\ M_{\text{ISM},\text{subhalo}} &\rightarrow \emptyset, \\ M_{\star,\text{host}} &\rightarrow M_{\star,\text{host}} + M_{\star,\text{subhalo}}, \\ M_{\star,\text{subhalo}} &\rightarrow \emptyset, \end{aligned} \quad (13)$$

where “ \emptyset ” indicates that a reservoir no longer exists. We do not include in our model any enhancement in the star formation rate after a merger (i.e. there is no explicit “starburst” mode), beyond any enhancement arising from the addition of new gas to a galaxy.

3 LIKELIHOOD

Perhaps the most important aspect of our calculation is the construction of a realistic likelihood function for our model. In this section we detail the calculation of this likelihood function.

3.1 Observed Stellar Mass Function and Covariance Matrix

We adopt the stellar mass function of galaxies at $z \approx 0.07$ measured from the Sloan Digital Sky Survey (SDSS) by Li & White (2009) as the single constraint on our model. Li & White (2009) report errors on their mass function which they derive from mock catalogs. However, to correctly evaluate the model likelihood we need to know the full covariance matrix of the observed mass function. We have developed a procedure, utilizing the framework of Smith (2012) to estimate the covariance matrix of this mass function.

The framework of Smith (2012) computes three contributions¹⁵ to the covariance matrix:

- (i) a Poisson term due to the finite number of galaxies in each bin;
- (ii) a large-scale structure term arising from structure in the Universe on scales comparable to and larger than the volume surveyed;
- (iii) a “halo” term arising from the fact that galaxies are found in associations (i.e. groups and clusters).

Calculating these three contributions for a given observational sample requires the following information (we refer the reader to Smith (2012) for a full explanation):

- (i) the survey geometry (i.e. the angular mask of the survey);
- (ii) the depth of the survey as a function of galaxy mass;
- (iii) knowledge of how galaxies populate dark matter halos, the so-called HOD.

For the angular mask, we make use of the catalog of random points within the survey footprint provided by the NYU-VAGC¹⁶ (Blanton et al. 2005; see also Adelman-McCarthy et al. 2008; Padmanabhan et al. 2008). Li & White (2009) consider only the main, contiguous region and so we keep only those points which satisfy $\text{RA} > 100^\circ$, $\text{RA} < 300^\circ$, and $\text{RA} < 247^\circ$ or $\delta < 51^\circ$. When the survey window function is needed, these points are used to determine which elements of a 3D grid fall within the window function.

To estimate the depth of the Li & White (2009) sample as a function of galaxy stellar mass we make use of semi-analytic models in the Millennium Database. Specifically, we use the SAM of De Lucia & Blaizot (2007; specifically the `millimil..DeLucia2006a` and `millimil..DeLucia2006a_sdss2mass` tables in the Millennium Database). For each snapshot in the database, we extract the stellar masses and observed-frame SDSS r-band absolute magnitudes (including dust extinction), and determine the median absolute magnitude as a function of stellar mass. Using the limiting apparent magnitude of the Li & White (2009) sample, $r = 17.6$, we infer the corresponding absolute magnitude at each redshift and, using our derived absolute magnitude–stellar mass relation, infer the corresponding stellar mass.

The end result of this procedure is the limiting stellar mass as a function of redshift, accounting for k-corrections, evolution, and the effects of dust. Figure 1 shows the resulting relation between stellar mass and the maximum redshift at which such a galaxy would be included in the sample. Points indicate measurements from the SAM, while the line shows a polynomial fit:

$$\begin{aligned} z(M_\star) = & -5.950 + 2.638m - 0.4211m^2 \\ & + 2.852 \times 10^{-2}m^3 - 6.783 \times 10^{-4}m^4, \end{aligned} \quad (14)$$

where $m = \log_{10}(M_\star/M_\odot)$. We use this polynomial fit to determine the depth of the sample as a function of stellar mass. We adopt a solid angle of 2.1901993 sr (Percival et al. 2007) for the sample.

Computing the large-scale structure contribution to the covariance function requires integration of the non-linear matter power spectrum over the Fourier transform of the survey window function. We use the method of Peacock & Dodds (1996) to determine the non-linear matter power spectrum, because of its simplicity and speed. We have checked that using a more accurate non-linear matter power spectrum (e.g. Lawrence et al. 2010) makes negligible difference to our results.

To find a suitable HOD to describe the galaxies in the

¹⁵ There are, of course, additional random and systematic uncertainties in stellar mass estimates. These will be addressed in §3.2.1 and §3.2.2 respectively.

¹⁶ Specifically, http://sdss.physics.nyu.edu/lss/dr72/random/lss_random-0.dr72.dat.

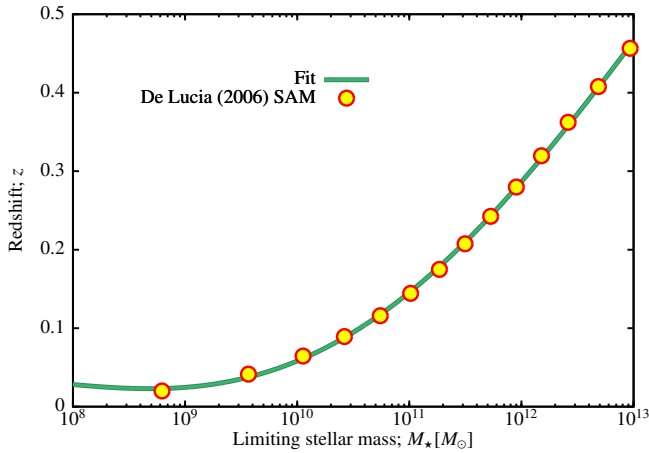


Figure 1. The maximum redshift at which a galaxy of given stellar mass can be detected in the sample of Li & White (2009). Points show the results obtained using the De Lucia & Blaizot (2007) model from the Millennium Database, while the lines shows a polynomial fit to these results (given in eqn. 14).

Li & White (2009) sample we adopt the model of Behroozi et al. (2010)¹⁷. This is an 11 parameter model which describes separately the numbers of satellite and central galaxies occupying a halo of given mass—the reader is referred to Behroozi et al. (2010) for a complete description of the functional form of this parametric HOD (see also Appendix B).

To reproduce the mass function of Li & White (2009), $\phi^{(\text{observed})}$, using this HOD we use the Bayesian Inference Engine (BIE) (Weinberg 2013) to constrain the HOD parameters. Details of this procedure are given in Appendix A. We use a likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \Delta \cdot \mathcal{C}^{-1} \cdot \Delta^T - \frac{N}{2} \ln(2\pi) - \frac{\ln |\mathcal{C}|}{2}, \quad (15)$$

where N is the number of bins in the mass function, \mathcal{C} is the covariance matrix of the observed mass function, and $\Delta_i = \phi_i^{(\text{HOD})} - \phi_i^{(\text{observed})}$, where $\phi^{(\text{HOD})}$ is the mass function computed from our HOD model. Of course, it is precisely this covariance matrix, \mathcal{C} , that we are trying to compute. We therefore adopt an iterative approach as follows:

- (i) make an initial estimate of the covariance matrix, assuming that only Poisson errors contribute (the covariance matrix is therefore diagonal, and the terms are easily computed from the measured mass function and the survey volume as a function of stellar mass);
- (ii) find the maximum likelihood parameters of the HOD given the observed mass function and the current estimate of the covariance matrix;
- (iii) using this HOD and the framework of Smith (2012),

¹⁷ One might ask, why not simply use the HOD from the SAM in this calculation of the data covariance matrix? This would indeed be the ideal solution, and one which we intend to investigate in future. For now, such an approach is practically impossible because it requires iteratively constraining the SAM to the data and updating the data covariance matrix on each iteration. This is too slow to be practical. However, planned improvements in the speed of data covariance matrix construction may allow this approach to be applied in future.

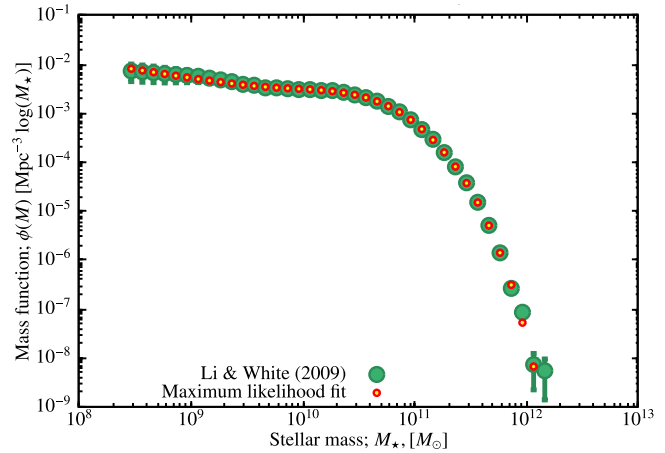


Figure 2. The maximum likelihood mass function obtained from our parametric HOD (yellow points), compared to the observed stellar mass function of Li & White (2009; green points).

compute a new estimate of the covariance matrix, including all three contributions;

- (iv) repeat steps 2 and 3 until convergence in the covariance matrix is achieved.

In practice we find that this procedure leads to an HOD and covariance matrix which oscillate between two states in successive iterations. The differences in the covariance matrix are relatively small however, so we choose to conservatively adopt the covariance matrix with the larger values. In future, adding additional constraints to the HOD (as described below) should help mitigate this problem.

The resulting maximum likelihood mass function is shown in Fig. 2, clearly illustrating that this parametric HOD can produce an excellent match to the observed mass function. The resulting correlation matrix is shown in Fig. 3. As expected, at the higher masses the correlation matrix is dominated by the on-diagonal terms—arising from the Poisson fluctuations in the number of galaxies due to the scarcity of these massive systems. At lower masses the matrix has significant off-diagonal amplitude, indicating strong correlations between nearby bins, arising from both large-scale structure and halo contributions to the covariance. This structure significantly weakens the constraint arising from the low-mass end of the mass function. Also noticeable are regions of enhanced correlation peaking at roughly $(4 \times 10^{11} M_\odot, 5 \times 10^{10} M_\odot)$. These arise from the “halo” term, describing the fact that galaxies are found in groups and clusters.

Li & White (2009) estimated errors on their mass function by generating 20 mock catalogs from the $z = 0$ Millennium Database (utilizing the SAM of Croton et al. 2006), and then measuring the variance of the resulting model stellar mass functions. This method should account for all of the sources of variance considered in our method. They report only the variance in each bin (i.e. the diagonal elements of the covariance matrix). We therefore compare the root-variance of the diagonal elements of our matrix with the errors reported by Li & White (2009). We find that our errors exceed those reported by Li & White (2009) in the lowest mass bins by a factor of approximately 3. In this regime, the variance is dominated by the “halo” term, suggesting that

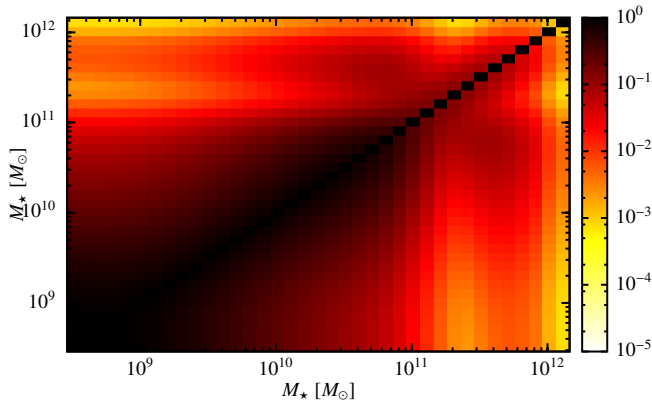


Figure 3. The correlation matrix of the observed galaxy stellar mass function of Li & White (2009). Color indicates the strength of correlation between bins, according to the scale shown on the right.

this difference lies in the relative distribution of galaxies between halos in the two models. For the highest mass bins, our estimates are a factor 2–3 smaller than those of Li & White (2009). Both approaches to estimating the errors on the mass function rely on the assumption of an underlying model. However, for the highest mass bins the errors should be dominated by the Poisson term, and therefore depend only on the measured mass function and the survey volume. Figure 4 shows the Croton et al. (2006) stellar mass function overlaid on the Li & White (2009) mass function. At high masses the Croton et al. (2006) model mass function exceeds that measured by Li & White (2009). This will lead to an overestimation of Poisson errors in the mass function in this regime. Our approach has the advantage that the underlying model used to determine the covariance is constrained to match the measured mass function to high precision. In any case, this illustrates one difficulty in determining covariances using models—the covariance is only as reliable as the model. As models are refined this should become less of a problem.

These techniques for covariance matrix estimation are applicable to mass functions and luminosity functions at any redshift, providing the survey geometry and depth are known. The only significant assumption is that the HOD model is a good description of the true galaxy population¹⁸. Here, we have self-calibrated the HOD to match the mass function being studied. It is possible, however, to introduce additional constraints on the HOD (e.g. the galaxy correlation function, or the evolution of the mass function with redshift) which would place tighter constraints on its parameters, or, possibly, rule it out as a viable description of the galaxy population. In the latter case, a more elaborate model would then be required. Indeed, given a complete galaxy formation model of the type studied in this work, it is possible to predict the HOD directly from that model. In future, once our models are more tightly constrained, it may be possible to use them to predict the HOD directly and use

¹⁸ Since the parameters of this HOD model are uncertain (as is the survey depth to some degree) we should arguably include this uncertainty when constructing our likelihood function. This point is discussed further at the end of Appendix A.

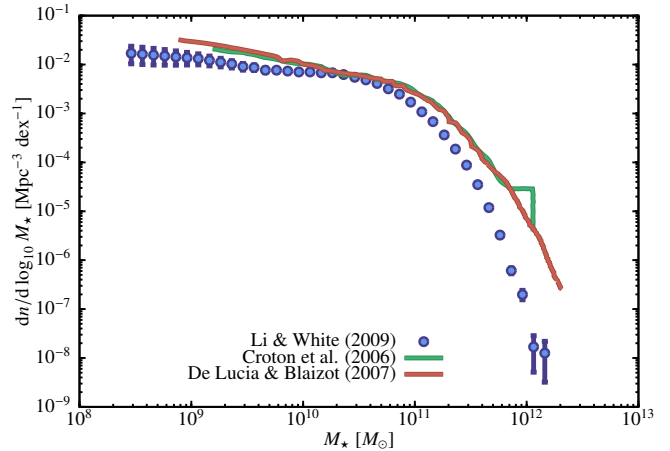


Figure 4. Comparison of three galaxy stellar mass functions. Blue points show the observationally determined stellar mass function from Li & White (2009). The green line shows the model stellar mass function of Croton et al. (2006; measured from Fig. 3 of Mutch et al. 2013), while the red lines shows the model stellar mass function of De Lucia & Blaizot (2007; measured from Fig. 7 of Kitzbichler & White 2007) which is a model very similar to that of Croton et al. (2006).

this to compute the covariance matrices of new observational datasets.

3.2 Errors in Stellar Mass Estimates

Stellar masses of galaxies are inferred from broadband photometry by fitting template spectral energy distributions (SEDs) (see Taylor et al. (2011) for a recent discussion and application of this methodology). This process includes many uncertainties, both random and systematic, due to photometric errors, stellar population synthesis, dust modeling, assumptions about possible star formation histories, and the initial mass function (IMF). To correctly compare model and data we must take into account these errors. In our analysis we follow the approach of Behroozi et al. (2010).

3.2.1 Random Errors

We assume that measured stellar masses in the SDSS have random errors distributed as a Gaussian in $\log_{10} M_*$ with a variance of $\sigma_*^2 = 0.07^2$ (Behroozi et al. 2010). Therefore, when constructing the stellar mass function from a model galaxy population we convolve the mass function with a Gaussian of this variance.

3.2.2 Systematic Errors

Estimated stellar masses may also be systematically biased due to the limitations of the template fitting procedure. A detailed discussion of these sources of systematic error can be found in Behroozi et al. (2010). We follow Behroozi et al. (2010) and assume that the systematic error in stellar mass can be modeled using:

$$\log_{10}(M_*) \rightarrow \log_{10}(M_*) + \mu + \kappa \log_{10} \left(\frac{M_*}{10^{11.3} M_\odot} \right). \quad (16)$$

Here, μ and κ are unknown coefficients. We treat these as free parameters in our MCMC analysis, and will marginalize over these parameters when determining the constraints on other parameters of the model. For priors on μ and κ , we follow Behroozi et al. (2010) and assume normal distributions with zero mean and dispersions of 0.27 and 0.1 respectively, truncated beyond ± 0.4 and ± 0.15 respectively.

Note that we are not fully including the IMF assumed in estimating SDSS stellar masses as a source of systematic uncertainty here. To first order, the choice of IMF leads to a relatively simple uniform shift in $\log M_*$. As such, if a different IMF were assumed a simple offset in the parameters of our model would be required to accommodate this change. However, in detail changes in the IMF will lead to changes in galaxy colours, potentially in a way not captured by our simple model for stellar mass systematics. Furthermore, as we do not incorporate physics related to stellar evolution (such as metal enrichment, or calculation of stellar population luminosities) in our simple model, the IMF need not be specified in our model. We intend to explore these points in greater detail in subsequent papers in this series.

3.3 Systematic Errors in the Halo Mass Function

Although the halo mass function has been determined to very high precision from cosmological N-body simulations (see, for example, Tinker et al. 2008) for specific halo finding algorithms, there remain significant systematic uncertainties in this function arising from the choice of halo finding algorithm. For example, Knebe et al. (2011) report relative differences of up to 30% in the halo mass function from a cross-comparison of 17 different halo finding algorithms. To account for this systematic uncertainty in the halo mass function, $n(M)$, we introduce a simple model of the systematic error in the halo mass function arising from the choice of halo finding algorithm. Specifically,

$$n(M) \rightarrow n(M)[1 + \alpha + \beta \log_{10}(M_{\text{halo}}/10^{12}M_{\odot})] \quad (17)$$

where α and β are nuisance parameters to be included in our MCMC analysis and marginalized over.

To determine priors on α and β , we make use of the relative residuals in halo mass functions determined by 14 halo finding algorithms reported in the lower panel of Fig. 17 of Knebe et al. (2011). We fit eqn. (17) to each line in that figure to determine the best-fit values of α and β , and then measure the means, $\langle \alpha \rangle$ and $\langle \beta \rangle$, and covariance, \mathcal{C} , of these parameters. We find

$$\begin{aligned} \langle \alpha \rangle &= -0.0198, \\ \langle \beta \rangle &= +0.0099, \\ \mathcal{C} &= \begin{pmatrix} 0.0091 & -0.0047 \\ -0.0047 & 0.0032 \end{pmatrix}. \end{aligned} \quad (18)$$

The means are consistent with zero given the variances and sample size. This is expected as the residuals in the lower panel of Fig. 17 of Knebe et al. (2011) were defined relative to the mean halo mass function from the 14 halo finding algorithms considered. To incorporate this multi-dimensional Gaussian prior into our likelihood function we define two new parameters H_1 and H_2 which each have standard nor-

mal priors construct α and β as:

$$\begin{aligned} \alpha &= -0.01979 + 0.09557H_1 \\ \beta &= +0.00986 - 0.04963H_1 + 0.02724H_2. \end{aligned} \quad (19)$$

3.4 Construction of Model Mass Function

To construct a model mass function, we generate a set of merger trees, and solve the system of ODEs and rules given in §2 to determine the stellar masses of model galaxies at $z = 0.07$. For each galaxy, the weight of its merger tree and the stellar mass are shifted to the cosmological model assumed in the analysis of the data¹⁹. Then the weight of the merger tree (eqn. 2) is distributed over a Gaussian of variance σ_*^2 in $\log_{10} M_*$, centered on the stellar mass of the galaxy shifted for systematic errors according to eqn. (16). This distribution is accumulated into bins which match those of the measured stellar mass function. That is:

$$\phi_m = \sum_{j=1}^{N_t} w_j \sum_{k=1}^{N_j} f_m(M'_*), \quad (20)$$

where ϕ_m is the mass function in the m^{th} bin, N_t is the number of merger tree realizations, N_j is the number of galaxies formed in tree j , $f_m(x)$ is the fraction of a lognormal of variance σ_*^2 and mean x that lies within the bounds of the m^{th} bin of the mass function, and M'_* is the stellar mass of the model galaxy with systematic errors applied according to eqn. (16). The covariance of the model mass function, $\mathcal{C}^{(\text{model})}$, which arises from the Monte Carlo nature of our model, is estimated as described in Appendix D, where we show that this model accurately reproduces the measured covariance of the model.

3.5 Model Discrepancy

Our model is of course approximate. As such, even if the correct model parameters were known with infinite precision it would still not precisely reproduce the properties of observed galaxies. To mitigate this fact, we attempt to model the “model discrepancies” which arise as a consequence of this approximate nature.

We assume that model discrepancy can be approximated by a combination of systematic shifts in stellar mass and in abundance (or by inflation of the prior on systematic shifts in stellar mass and inflation of the model covariance matrix). Specifically, we model a systematic shift in stellar masses due to model discrepancy using the model given by eqn. (16). Once the best-fitting values of μ and κ have been found the remaining discrepancy is described as an offset plus multi-variate Gaussian covariance, $\mathcal{C}^{(\text{discrepancy})}$, in the mass function ϕ . In general, we adopt a model in which

$$\phi_i \rightarrow \zeta_i + \eta_i \phi_i, \quad (21)$$

where subscript i refers to the i^{th} bin of the mass function,

¹⁹ That is, stellar masses are multiplied by the ratio of the square of the luminosity distance to $z = 0.07$ in the data and model cosmologies, while the merger tree weight is multiplied by the ratio of $d_A^2 dr_c/dz$ in those two cosmologies, where d_A is angular diameter distance and r_c is comoving distance, and all terms are evaluated at $z = 0.07$.

ϕ , and ζ and η are coefficients. That is, we consider both constant offset and multiplicative offset contributions to the discrepancy.

We will estimate model discrepancies using our model itself. Therefore, we must adopt some set of model parameters to use for this purpose. As we do not know the optimal set of parameters in advance (discerning this optimal set being the purpose of our study) we instead adopt a set of “*a priori*” parameters, described in Appendix C, for this purpose. This approach will be valid providing that the *a priori* model parameters are sufficiently close to the maximum of the model posterior found by our MCMC analysis. We will address this point in §4.3.4.

3.5.1 Fixed Orbital Parameters

As noted in §2.3.1, when computing the merging timescale of dark matter subhalos, we assign each subhalo orbital parameters equal to the mode of the measured cosmological distribution. That is, we ignore the scatter in orbital parameters. We do this to allow our models to converge more rapidly and reduce the CPU time required for calculations. This is, therefore, an approximation which we should account for using a model discrepancy term.

To compute the model discrepancy, we run two calculations using our *a priori* model: one with fixed orbital parameters, and a second with orbital parameters drawn from the full cosmological distribution. We then define

$$\eta_i = \phi_i^{(\text{variable})} / \phi_i^{(\text{fixed})}, \quad (22)$$

where $\phi_i^{(\text{variable})}$ and $\phi_i^{(\text{fixed})}$ are the stellar mass functions in the i^{th} in the model with variable and fixed orbital parameters respectively. The covariance, $\mathcal{C}^{(\text{discrepancy})}$, is set to the sum of the covariances of the two models²⁰. We assume that the distribution of η can be modeled as a multivariate log-normal distribution with this covariance. We find that η_i is $100 \pm 1\text{--}2\%$.

3.5.2 (Lack of) Scatter in Merging Times

Jiang et al. (2008) report a scatter between merger times computed from simulations and their fitting formula for merger timescale. This scatter is well described by a log-normal with width $\sigma \approx 0.4$. In the models run in our MCMC analysis we ignore this scatter to reduce the CPU time needed for models to converge. We account for this approximation via a model discrepancy term.

To compute the model discrepancy, we run two calculations using our *a priori* model: one with no scatter in merger times, and a second with $\sigma = 0.4$ as recommended by Jiang et al. (2008). We then define

$$\eta_i = \phi_i^{\sigma=0.4} / \phi_i^{\sigma=0.0}, \quad (23)$$

where $\phi_i^{\sigma=0.4}$ and $\phi_i^{\sigma=0.0}$ are the stellar mass functions in the i^{th} in the model with and without scatter respectively. The

²⁰ For all model discrepancy evaluations we run models much larger than those used in our MCMC analysis. Therefore, their covariance is a sub-dominant contribution to the total covariance, except where noted. Nevertheless, we include it in our estimate of the total covariance.

covariance, $\mathcal{C}^{(\text{discrepancy})}$, is set to the sum of the covariances of the two models. We assume that the distribution of η can be modeled as a multivariate log-normal distribution with this covariance. We find that η_i is $100 \pm 1\text{--}5\%$ with the effect being largest in the highest mass bins.

3.5.3 Monte Carlo-Generated Merger Trees

Our merger trees are generated using a modification of the extended Press-Schechter formalism (Parkinson et al. 2008). As such, they do not match perfectly the statistics of merger trees drawn from N-body simulations, although they are very close (see Appendix E). We note that we are specifically *not* assuming that the N-body trees represent the “correct” answer here. Merger tree construction algorithms give significantly different results (Srisawat et al. 2013), and the relatively small number of snapshots in the Millennium Simulation may limit the efficacy with which merger tree building algorithms can perform. Furthermore, Jiang & van den Bosch (2013) note that the algorithm of Parkinson et al. (2008) produces merger trees that differ from those extracted from N-body simulations by no more than the variation in N-body merger trees given different analysis methods. As such, we will treat this discrepancy as a source of covariance only (i.e. with no systematic offset) to reflect the fact that we do not know which type of tree (Monte Carlo or N-body) is more correct.

To quantify this model discrepancy we run our *a priori* model on merger trees extracted from all 512 sub-volumes of the Millennium Simulation (Springel et al. 2005). We adjust the cosmological parameters and power spectrum assumed in our *a priori* model to precisely match those used in the Millennium Simulation. We force the merger trees to be monotonically growing in mass (see Appendix E). We then run the same model (i.e. still using the Millennium Simulation cosmology and power spectrum) using Monte Carlo merger trees, with the root masses of the trees at $z = 0$ precisely matched to the masses of $z = 0$ Millennium Simulation halos to remove any effects of variance due to finite volume. In this case, we re-sample the merger trees onto the set of times corresponding to the Millennium Simulation snapshots (for details of this re-sampling process, see Benson et al. 2012).

We parameterize the difference between the two mass functions using the systematic mass model of eqn. (16), finding best-fit values of $\mu = -0.1025$, and $\kappa = -0.0225$. As there is no reason to assume that either Monte Carlo or N-body merger trees are “better”, we do not shift the means of the priors on these parameters, but instead add the above values in quadrature to the variance of their Gaussian priors.

After applying this systematic shift to the Monte Carlo models, we assume a contribution to the covariance matrix of

$$\mathcal{C}_{ij}^{(\text{discrepancy})} = [\phi_i^{(\text{N-body})} - \phi_i^{(\text{MC})}][\phi_j^{(\text{N-body})} - \phi_j^{(\text{MC})}], \quad (24)$$

where $\phi_i^{(\text{N-body})}$ and $\phi_i^{(\text{MC})}$ are the stellar mass functions in the i^{th} in the models using N-body and Monte Carlo merger trees respectively. For $\phi_i^{(\text{N-body})}$ we use the mass function estimated from the union of all 512 sub-volumes of the Millennium Simulation. We add to this covariance the sum of the internal covariances of the two models.

For the N-body merger trees, there is, additionally, some covariance arising from the finite volume of the Millennium Simulation. To estimate the total covariance in these models, including this finite volume source, we measure the covariance between the 512 sub-volumes of the Millennium Simulation. We begin by measuring the covariance between individual sub-volumes. We then proceed to pair sub-volumes and measure the covariance between pairs of sub-volumes. We repeat this process, doubling the number of sub-volumes combined each time, until we are combining 64 sub-volumes. From this we measure the dependence of the covariance matrix on the number of sub-volumes combined. We extrapolate this dependence to 512 sub-volumes to estimate the covariance in the full Millennium Simulation.

Over the power-law range of the mass function, we find that $\phi_i^{(N\text{-body})}$ and $\phi_i^{(MC)}$ differ by typically $\sim 1\%$. At the exponential cut-off, the N-body model predicts a significantly different mass function from the Monte Carlo merger trees, which makes the ratio of $\phi_i^{(N\text{-body})}$ to $\phi_i^{(MC)}$ first decline to around 0.3 before beginning to diverge to large values in the final few mass bins. This discrepancy is therefore large, but is nevertheless included. This is an aspect of the model discrepancy which warrants further study in future. In any case, the covariance introduced by this model discrepancy becomes large in the final few bins of the mass function, which will cause them to make only a small contribution to our likelihood.

3.5.4 Variable Mass Resolution in Merger Trees

To permit rapid calculation of the model expectation, we use a variable mass resolution in our merger trees as described in eqn. (1). This is an approximation which affects the model estimate of the galaxy stellar mass function, which we account for using a model discrepancy term.

To compute the model discrepancy, we run two calculations using our *a priori* model: one with variable mass resolution as described in eqn. (1), and a second with fixed mass resolution ($M_{\text{res,min}} = 5 \times 10^9 M_\odot, f_0 = 0$). We assume that this discrepancy can be described by a constant offset in abundance (accounting for the missing halos in each tree) and a (possibly mass-dependent) systematic offset in stellar mass. We find that the systematic difference in these mass functions is well described by the systematic mass model of eqn. (16) with $\mu = -0.0125$, $\kappa = 0.165$. As such, we set the means of the priors on these parameters to these values. After applying this shift to the fixed mass resolution model, we then define

$$\eta_i = \phi_i^{(\text{fixed})} / \phi_i^{(\text{variable})}, \quad (25)$$

where $\phi_i^{(\text{fixed})}$ and $\phi_i^{(\text{variable})}$ are the stellar mass functions in the i^{th} in the model with fixed and variable mass resolution respectively. The covariance, $\mathcal{C}^{(\text{discrepancy})}$, is set to the sum of the covariances of the two models. We assume that the distribution of η can be modeled as a multivariate log-normal distribution with this covariance.

Note that we define the abundance discrepancy in all bins as equal to that in the lowest mass bin consistent with our assumption of a constant offset in abundance. This avoids the discrepancy term becoming uncomfortably large in the high mass bins. In these bins it is our opinion that the simple mass discrepancy model is insufficient

to capture the details of this model discrepancy. Therefore, instead of applying a large correction, we include additional covariance to account for this failing of our discrepancy model. Specifically, any remaining offset, η' , in the ratio $\phi_i^{(\text{fixed})} / \phi_i^{(\text{variable})}$ after applying the η_i correction described above is used to add to the covariance matrix, $\mathcal{C}_{ij}^{(\text{discrepancy})} \rightarrow \mathcal{C}_{ij}^{(\text{discrepancy})} + \phi_i \phi_j \log(\eta'_i) \log(\eta'_j)$.

We find that η_i is approximately 95%. The additional covariance due to the limitations of our mass systematic model becomes large in the six highest mass bins, reducing the constraining power of the data in this regime. This is clearly a very large correction, but currently necessary due to practical limitations.

3.5.5 Cosmology-independent Merger Trees

As discussed in §2.1.1, we use a single set of merger trees in all model evaluations (to avoid introducing discontinuities in the likelihood surface). This means that cosmological dependence of merger tree structure is lost from the model. To account for this fact we introduce a model discrepancy term which increases the model covariance. To compute this term we generate a large number of model realizations with cosmological parameters and halo mass systematic parameters drawn from their respective priors (see §3.3 and §3.7.2). We run one such set of models using a fixed set of merger trees (i.e. with trees generated using the maximum likelihood cosmological and systematics parameters), and a second set in which we generate a new set of trees for each model evaluation. We then construct the covariance matrix of the stellar mass function from each set of models. The difference between the covariance of the set with cosmology-dependent trees and those with cosmology-independent trees is taken as our model discrepancy term.

3.6 Final Likelihood

We assume that both the model and observed mass functions can be described as multivariate normal distributions, and so the model likelihood is finally evaluated as

$$\ln \mathcal{L} = -\frac{1}{2} \Delta \mathcal{C}^{-1} \Delta^T - \frac{N}{2} \ln(2\pi) - \frac{\ln |\mathcal{C}|}{2}, \quad (26)$$

where N is the number of bins in the mass function, the covariance matrix is given by

$$\mathcal{C} = \mathcal{C}^{(\text{observed})} + \mathcal{C}^{(\text{model})} + \mathcal{C}^{(\text{discrepancy})}, \quad (27)$$

where $\Delta_i = \zeta_i + \eta_i \phi_i^{(\text{model})} + \epsilon_i - \phi_i^{(\text{observed})}$ (recalling that ζ_i and η_i describe model discrepancies—see eqn. 21).

3.7 MCMC Procedure

In the following sections we describe the MCMC procedure used to find constraints on model parameters.

3.7.1 Parameter Priors

Table 1 summarizes the priors that we adopt for our model parameters. In this table, $U(x_l, x_u)$ indicates a uniform prior on the parameter over the range $[x_l, x_u]$ and zero outside of

that range. $U_{\text{in}}(x_1, x_h)$ indicates a uniform prior in the logarithm of the parameter over the range $[x_1, x_u]$ and zero outside of that range. $N(\mu, \sigma^2, x_1, x_h)$ indicates a normal prior with mean μ , variance σ^2 , and truncated to zero outside the range $[x_1, x_u]$.

Priors on composite cosmological parameters, $C_{1...6}$, and halo mass function systematics, $H_{1...2}$, are standard normals by construction (see §3.7.2 and §3.3 respectively), while the priors on the parameters of our model for systematics in stellar masses are described in §3.2.2. For all other parameters we adopt relatively uninformative priors, as follows:

α_{cool} We expect cooling to be more rapid in the past (densities are higher so cooling rates will be boosted and freefall timescales will be shorter). We therefore allow this parameter to range from a very negative value of -10 to $+1$ (to allow for the possibility that cooling actually becomes slightly less efficient at higher redshifts).

β_{cool} This parameter controls the sharpness cut-off in cooling efficiency. Given that the mass function cuts off sharply at high mass we expect this parameter should be positive and probably greater than unity. We set a range 0 to 8 .

$\Delta \log_{10} \mathcal{M}_{\text{cool}}$ Given the sharpness of the cut-off in cooling efficiency, we expect that this parameter will have to be relatively small, of order unity. We therefore limit it to the range 0.3 to 3.0 .

τ_{cool} Timescales for cooling (or, more precisely, for infall) at the present day should be no shorter than typical halo dynamical times (which are of order 1 Gyr), but could be much longer if heating processes (perhaps from outflows or active galactic nuclei (AGN)) are at work. We therefore adopt a uniform prior in the logarithm of this parameter between 1 Gyr and 3×10^3 Gyr.

$\mathcal{M}_{\text{cool}}$ This parameter is intended to model the mass scale at which cooling transitions from efficient to inefficient. Based on studies such as Benson et al. (2003) we expect this scale must lie slightly above the halo mass associated with L_* galaxies. If the cut-off in efficient cooling is associated with the transition from cold-mode to hot-mode cooling then we would expect $\mathcal{M}_{\text{cool}}$ to be of order a few times $10^{11} M_{\odot}$ (Kereš et al. 2005). To be conservative we therefore limit this parameter to the range 10^{10} to $3 \times 10^{13} M_{\odot}$.

α_* We do not have a strong intuition for the expected range for this parameter, nor a good way to infer it more directly from other data. Therefore, our prior is based on our experience of trial runs of our model. It will be important therefore to check whether the posterior on this parameter lies well within the prior or not.

β_* As for α_* our prior here is based on trial runs of our model.

$\tau_{*,\text{min}}$ The minimum timescale for star formation is limited to be sufficiently short that it will not affect the model results, while ensuring that model integration does not become unreasonably slow.

τ_* Typical specific star formation rates for star forming galaxies are of order 0.1 to 1.0 Gyr^{-1} at $z \approx 0$ (Brinchmann et al. 2004). Assuming a typical gas fraction of order 10% for galaxies with rotation speeds of approximately 200 km/s at the present day, this translate to a star formation timescale of around 0.1 to 1.0 Gyr. As such, we set the prior on the star formation timescale to lie in the range 0.03 to 1 Gyr.

α_{wind} It is not immediately obvious how wind mass loadings should scale with redshift, but we expect that outflows may be less strong at higher redshifts since galaxies will be denser (and so more energy may be radiated away rather than coupled to an outflow). Therefore, we allow a broad range for this parameter extending from very negative values to slightly above zero.

β_{wind} We expect outflows to be stronger in systems of lower virial velocity, which suggests this parameter should be negative. We therefore allow it to range from a very negative value (more negative than any plausible physical outflow model) to zero.

f_{wind} Martin et al. (2012) report mass-loading factors of order unity. Since we wish to also explore the possibility of models with minimal outflow we let this parameter range from a very small value to unity.

$\tau_{\text{wind,min}}$ The minimum timescale for outflows is limited to be sufficiently short that it will not affect the model results, while ensuring that model integration does not become unreasonably slow.

f_{df} Merging timescales should be well calibrated through our use of the Jiang et al. (2008) fitting formula. However, we allow for the possibility for variations in this timescale (due to approximations in how merging is defined, effects of baryonic physics on merging, etc.) by adopting a prior that is uniform in the logarithm of this parameter between 0.1 and 10 .

V_{reion} The effects of reionization on the accretion of gas into halos has been well studied. For example, Okamoto et al. (2008) find that accretion is suppressed in systems below a characteristic velocity of roughly 25 km/s . Font et al. (2011) explore the same, simple truncation model for the effects of reionization that we adopt in this work. They find that the results of Okamoto et al. (2008) can be reproduced in such a model with $V_{\text{reion}} = 34 \text{ km/s}$. Previous studies (Gnedin 2000) had reported much stronger effects from reionization. To explore the full range of possibilities we therefore, we adopt a prior that is uniform between 30 and 60 km/s .

3.7.2 Cosmological Parameters

We include six cosmological parameters in our MCMC analysis. Even though the values of these parameters are well constrained by combinations of cosmic microwave background, supernovae, Cepheid, lensing, and baryon acoustic oscillation experiments (see Planck Collaboration et al. 2013, for a review) some uncertainty in their values remain (particularly for σ_8), and we wish to assess the impact of these uncertainties on the model posterior and any predictions from that posterior. Therefore, we make use of the WMAP 9-year data (Hinshaw et al. 2013) to define priors on the combination of parameters $(H_0, \Omega_M, \Omega_b, \sigma_8, n_s, \tau)$, and assume a flat universe such that $\Omega_{\Lambda} = 1 - \Omega_M$. Using the MCMC chains provided by the WMAP collaboration²¹

²¹ Specifically the chains combining constraints from WMAP-9, SPT (Keisler et al. 2011), ACT (Das et al. 2011), SNLS3 (Sullivan et al. 2011), BAO (Beutler et al. 2011; Padmanabhan et al. 2012; Anderson et al. 2012; Blake et al. 2012), and H_0 (Riess et al. 2011; Freedman et al. 2012):

Parameter	Units	Description	Prior
α_{cool}	–	Exponent of $1/(1+z)$ in cooling rate	$U(-10, +1)$
β_{cool}	–	Exponent of cut-off mass-scale for efficient cooling	$U(0, +8)$
$\Delta \log_{10} \mathcal{M}_{\text{cool}}$	–	Width of cut-off for efficient cooling	$U_{\text{ln}}(3 \times 10^{-1}, 3 \times 10^1)$
τ_{cool}	Gyr	Timescale for cooling	$U_{\text{ln}}(1, 3 \times 10^3)$
$\mathcal{M}_{\text{cool}}$	M_{\odot}	Cut-off mass-scale for efficient cooling	$U_{\text{ln}}(10^{10}, 3 \times 10^{13})$
α_{\star}	–	Exponent of $(1+z)$ in star formation rate	$U(0, +4)$
β_{\star}	–	Exponent of virial velocity in star formation rate	$U(-8, -2)$
$\tau_{\star, \text{min}}$	Gyr	Minimum timescale for star formation	$U_{\text{ln}}(10^{-5}, 10^{-1})$
τ_{\star}	Gyr	Timescale for star formation	$U_{\text{ln}}(3 \times 10^{-2}, 1)$
α_{wind}	–	Exponent of $(1+z)$ in wind outflow rate	$U(-14, +1)$
β_{wind}	–	Exponent of virial velocity in wind outflow rate	$U(-20, 0)$
f_{wind}	–	Normalization of wind outflow rate	$U_{\text{ln}}(10^{-5}, 1)$
$\tau_{\text{wind, min}}$	Gyr	Minimum timescale for wind outflow rate	$U_{\text{ln}}(10^{-3}, 10^{-1})$
f_{df}	–	Multiplier for subhalo merging times	$U_{\text{ln}}(10^{-1}, 10^{+1})$
V_{reion}	km/s	Velocity scale for suppression of accretion from the IGM	$U(30, 60)$
$C_{1\dots 6}$	–	Cosmological parameter combination	$N(0, 1, -\infty, +\infty)$
μ	–	Coefficient in stellar mass systematic (eqn. 16)	$N(-0.0125, 0.0833, -0.5, +0.5)$
κ	–	Coefficient in stellar mass systematic (eqn. 16)	$N(0.165, 0.0105, -0.32, +0.32)$
$H_{1\dots 2}$	–	Halo mass function systematic error parameter combination	$N(0, 1, -\infty, +\infty)$

Table 1. Priors on parameters. $U(x_1, x_h)$ is uniform in the parameter over the range $[x_1, x_h]$ and is zero outside of that range. $U_{\text{ln}}(x_1, x_h)$ is uniform in the logarithm of the parameter over the range x_1 – x_h and is zero outside of that range. $N(\mu, \sigma^2, x_1, x_h)$ is a normal distribution with mean μ , variance σ^2 , and truncated to zero outside the range $[x_1, x_h]$.

we measure the covariance matrix of these six parameters and model their prior as a multivariate normal distribution. Specifically, we define six meta-parameters, $C_1 \dots 6$, each with a standard normal prior and construct the cosmological parameters from these with the required covariances as follows:

$$H_0 = 69.55 + 0.4011C_1 - 0.5585C_2 + 0.003770C_3 + 0.3875C_4 \text{ km/s/Mpc}, \quad (28)$$

$$\Omega_M = \min([0.1371 + 0.0004757C_1 + 0.001892C_2] \times [H_0/100 \text{ km/s/Mpc}]^{-2}, 1), \quad (29)$$

$$\Omega_\Lambda = 1 - \Omega_M, \quad (30)$$

$$\Omega_b = (0.02225 + 0.0003305C_1) \times (H_0/100 \text{ km/s/Mpc})^{-2}, \quad (31)$$

$$\sigma_8 = 0.8178 + 0.003817C_1 + 0.007931C_2 + 0.01002C_3 + 0.001584C_4 + 0.0029312C_5 + 0.001727C_6, \quad (32)$$

$$n_s = 0.9616 + 0.005302C_1 - 0.001576C_2 + 0.0009843C_3 + 0.0006839C_4 + 0.005735C_5, \quad (33)$$

$$\tau = \max(0.08155 + 0.001718C_1 - 0.001858C_2 + 0.01179C_3, 0). \quad (34)$$

The redshift of reionization, z_{reion} , is then found by solving the equation

$$\tau = \int_0^{z_{\text{reion}}} n_e(z) \frac{dl}{dz} dz, \quad (35)$$

where $n_e(z)$ is the electron density at redshift z assuming a fully-ionized universe.

3.7.3 MCMC

We use the BIE to perform a differential evolution MCMC simulation (Terr Braak 2006) to constrain the parameters of our model. We use 288 parallel chains for our main calculation. The state of each chain is initialized by drawing at random from the prior distributions. At each step of the simulation a proposed state, S'_i , for each chain, i , is constructed by selecting at random (without replacement) two other chains, m and n , and finding

$$S'_i = S_i + \gamma(S_m - S_n) + \epsilon, \quad (36)$$

where γ is a parameter chosen to keep the acceptance rate of proposed states sufficiently high, and ϵ is a random vector each component of which is drawn from a Cauchy distribution with median zero and width parameter set equal to 0.1% of the width of the prior²² to ensure that the chains are positively recurrent. For a multivariate normal likelihood function in N dimensions the optimal value of γ is $\gamma_0 = 2.38/\sqrt{N}$ (Terr Braak 2006). The proposed state is accepted with probability P where

$$P = \begin{cases} 1 & \text{if } \mathcal{L}(S'_i) > \mathcal{L}(S_i), \\ \mathcal{L}(S'_i)/\mathcal{L}(S_i) & \text{otherwise.} \end{cases} \quad (37)$$

We begin our MCMC simulation at a high temperature of $T = 128$. That is, log-likelihoods are divided through by T to smooth the likelihood surface and make it easier for the chains to rapidly explore the surface. The simulation is allowed to progress until the chains have converged on the heated posterior distribution as judged by the Gelman-Rubin statistic, \hat{R} (Gelman & Rubin 1992), after outlier chains (identified using the Grubb's outlier test (Grubbs 1969; Stefansky 1972) with significance level $\alpha = 0.05$) have

²² The width of the prior is taken to be the maximum minus minimum values for uniform priors and the standard deviation (or the upper minus lower limit if that is larger) for normal priors.

Table 2. The sequence of temperatures and corresponding values of γ used in our MCMC simulations.

Temperature	γ
128	0.3365
32	0.3365
8	0.1743
2	0.1200
1	0.1000

been discarded. Specifically, we declare convergence when $\hat{R} = 1.2$. The final state of the chains is then used as the starting point for a new simulation with lower temperature. We repeat this process, reducing the temperature each time, until we reach $T = 1$. Once the chains have converged for $T = 1$ we let them run to generate a sufficiently large sampling of the converged, unheated posterior. We find that this gradual cooling is an effective way to achieve convergence onto the final posterior. Table 2 lists the temperatures and corresponding values of γ used.

4 RESULTS

Before beginning the MCMC algorithm, we check that the numerical parameters of our model are sufficient to give robust, converged results. The results of this numerical study are presented in Appendix C.

4.1 Parameter Constraints

Figure 5 shows 1-D and 2-D marginalized posterior probability distributions for all 25 parameters in our analysis, while Table 3 lists parameter values at various percentages of their 1-D marginalized cumulative posterior probability distributions. The posteriors of the “nuisance” parameters (i.e. those describing systematic uncertainties in stellar mass, the halo mass function, and cosmological parameters) are dominated by their priors (indicating that the data do not provide significant additional constraint on these quantities). While the majority of the model parameters are rather poorly constrained, a handful of them are strongly constrained by the observed stellar mass function.

Figure 6 shows enlarged versions of two cases where the posterior is well constrained. In the left hand panel we plot the marginalized posterior for β_* and $\Delta \log_{10} \mathcal{M}_{\text{cool}}$. These parameters are both well constrained by the data (in the sense that their posteriors are much narrower than their priors), and are uncorrelated. The right hand panels shows the joint posterior of $\Delta \log_{10} \mathcal{M}_{\text{cool}}$ and β_{cool} . This pair of parameters is jointly well constrained, but highly correlated.

Finally, the posteriors of the parameters α_* and β_* —for which we did not have good pre-existing knowledge from which to set priors—are peaked well inside the ranges allowed by their priors. As such, our choice of prior did not strongly affect the results²³.

²³ It is always possible that an island of high-likelihood exists outside the range of the priors that we set.

We will discuss implications of the posterior in §5.

4.2 Posterior Predictive Checks

As described by Lu et al. (2012), posterior predictive checks (PPCs) (Gilks 1995; Gelman et al. 2013) are a powerful means by which to check that the model family characterized by the posterior probability distribution is a viable description of the observed data. We follow the procedure of Lu et al. (2012) to assess the PPC, and the reader is directed to that work for a complete description. Specifically, we adopt a test-statistic similar to that of Lu et al. (2012; eqn. 6):

$$\mathcal{T}_l = \mathcal{T}(\mathbf{y}_l) = \Delta_l \cdot \mathcal{C}_{\mathbf{y}}^{-1} \cdot \Delta_l^T, \quad (38)$$

where $\Delta_l = \mathbf{y}_l - \bar{\mathbf{y}}$, \mathbf{y}_l is the l^{th} model mass function sampled from our converged MCMC chains, $\bar{\mathbf{y}}$ is the mean model mass function over the MCMC chains, and $\mathcal{C}_{\mathbf{y}}$ is the diagonalized covariance matrix²⁴ of the set of model results. We compute the same quantity for the observational data

$$\mathcal{T}' = \Delta' \cdot \mathcal{C}_{\mathbf{y}}^{-1} \cdot \Delta'^T, \quad (39)$$

where $\Delta' = \mathbf{y}' - \bar{\mathbf{y}}$, and \mathbf{y}' is the observed mass function. The Bayesian p -value is then

$$\hat{p}_B = \frac{1}{L} \sum_{l=1}^L I_{\mathcal{T}_l \geq \mathcal{T}'}. \quad (40)$$

The left-hand panel of Figure 7 shows the cumulative probability of the test statistic, \mathcal{T} , for the model (yellow line) and indicates the test statistic of the data by the green arrow, showing that \mathcal{T}' lies in the tail of the distribution of model test statistics ($\hat{p}_B = 0.045$). The model is therefore only a very marginally good description of the data (as measured by this particular test statistic). The right-hand panel of Figure 7 shows the observed stellar mass function (blue points) along with the distribution of model mass functions drawn from the posterior probability distribution. Specifically, we sample 10,000 points from the converged MCMC chains (excluding outlier chains) and evaluate the model mass function for each point. We then find the median of the distribution of model mass functions in each mass bin. This is shown by the yellow line. Additionally, shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3-“ σ ” respectively) of the distribution of model mass functions. The model clearly closely follows the data, but noticeably lies systematically below the data at the low-mass end (e.g. roughly 85% of models lie below the data points). The knee and high-mass end of the mass function are well-matched—the “bump” in the model mass function just below $10^{12} M_{\odot}$ arises from the small number of model merger trees used to construct the mass function in this regime.

²⁴ We use the diagonalized covariance matrix here as it is more robust—we find that using the full covariance matrix estimated from this sample of 10,000 model realizations can lead to imprecisions in the matrix inversion to get \mathcal{C}^{-1} leading to wildly varying test statistics. Future studies might avoid this problem (and so be able to employ the full covariance matrix) by running the MCMC chains for longer thereby allowing a much larger number of independent model realizations to be generated.

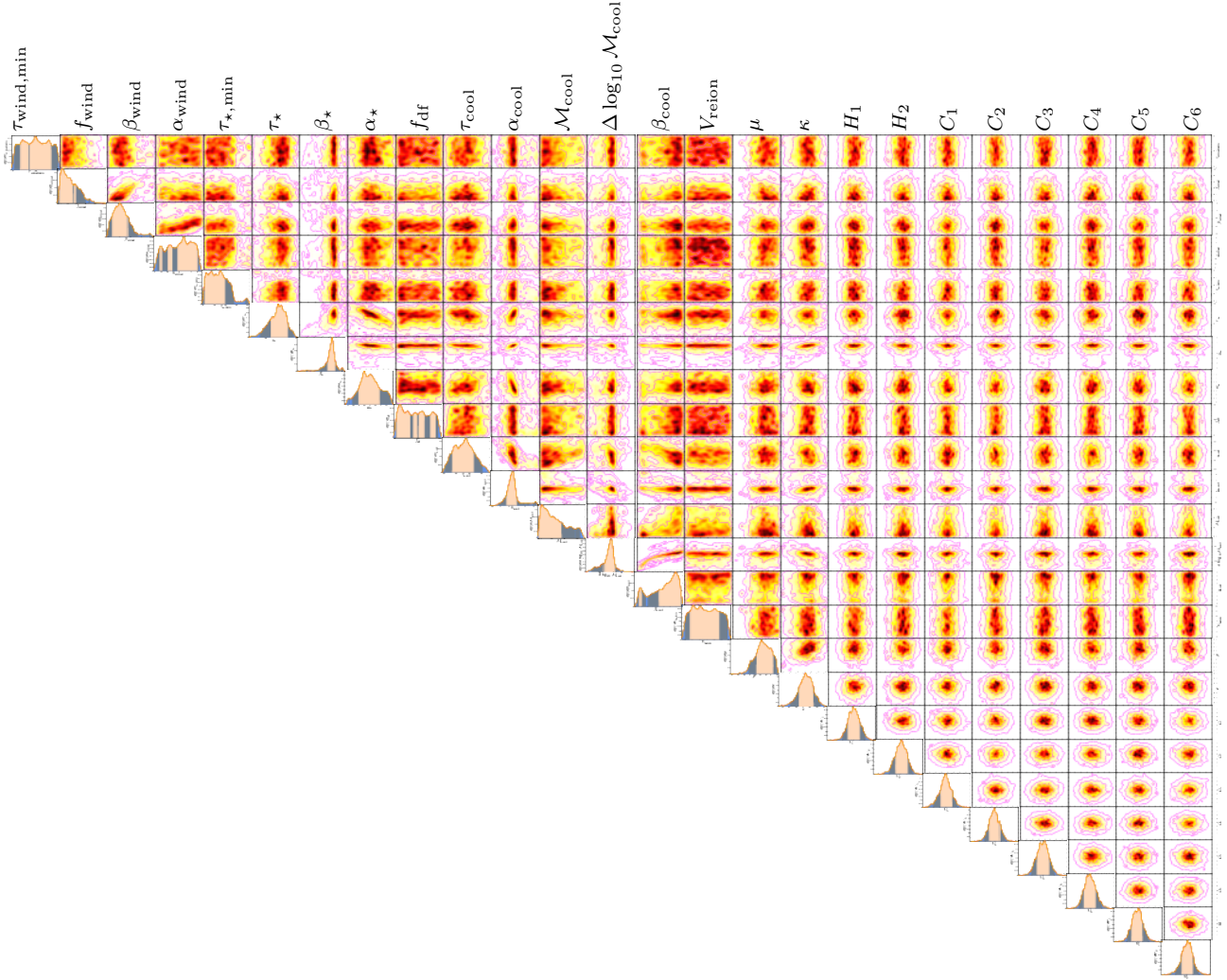


Figure 5. The posterior probability distribution of parameters in our model constrained to fit the observed stellar mass function of Li & White (2009). Panels on the diagonal show the marginalized posterior probability distribution for each individual parameter, while off-diagonal panels show the posterior probability distribution for pairs of parameters, marginalized over all other parameters.

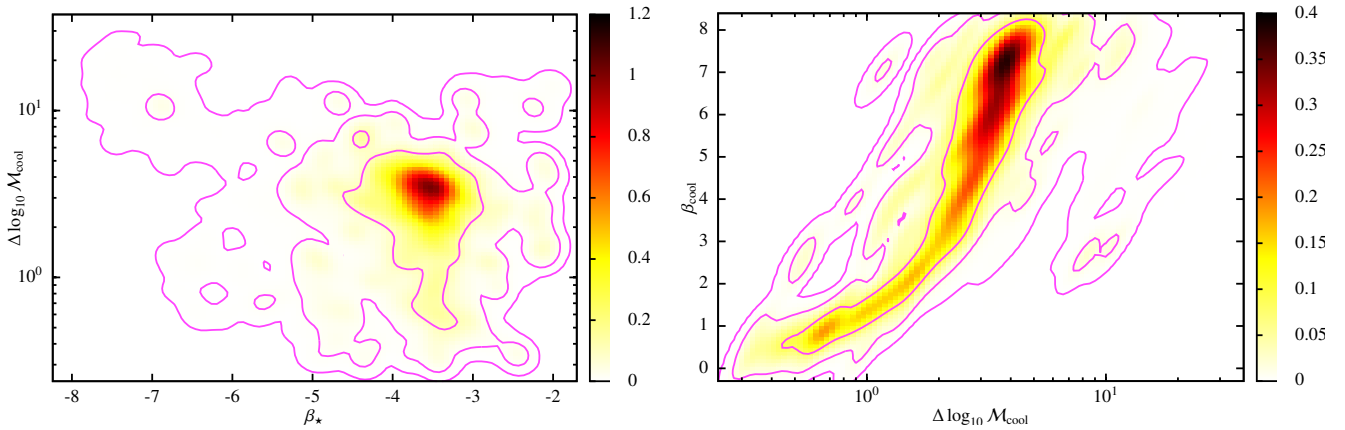


Figure 6. *Left panel:* The joint posterior probability distribution of β_* (the exponent of halo virial velocity appearing in the star formation rate; eqn. 6) and $\Delta \log_{10} M_{cool}$ (the width in the logarithm of halo mass of the cut-off in cooling efficiency; eqn. 5). *Right panel:* The joint posterior probability distribution of $\Delta \log_{10} M_{cool}$ and β_{cool} (the exponent of the cut-off factor in cooling efficiency; eqn. 5).

Table 3. Values of model parameters at the indicated percentages of their 1-D marginalized cumulative posterior probability distributions.

Parameter	Cumulative probability						
	0.27%	4.55%	31.73%	50.00%	68.27%	95.45%	99.73%
$\tau_{\text{wind,min}}$	0.845×10^{-3}	0.117×10^{-2}	0.417×10^{-2}	0.940×10^{-2}	0.0214	0.0781	0.111
f_{wind}	0.715×10^{-5}	0.117×10^{-4}	0.428×10^{-4}	0.120×10^{-3}	0.428×10^{-3}	0.0113	0.180
β_{wind}	-20.3	-19.2	-16.0	-14.4	-12.6	-6.09	-0.376
α_{wind}	-14.3	-13.0	-8.55	-5.72	-3.49	0.192	1.29
$\tau_{\star,\text{min}}$	0.746×10^{-5}	0.122×10^{-4}	0.615×10^{-4}	0.192×10^{-3}	0.556×10^{-3}	0.0107	0.105
τ_{\star}	0.0306	0.0673	0.173	0.242	0.321	0.611	0.884
β_{\star}	-7.20	-5.03	-3.90	-3.67	-3.48	-2.78	-2.07
α_{\star}	-0.0526	0.487	1.55	2.00	2.47	3.67	4.02
f_{df}	0.0836	0.114	0.344	0.830	1.94	7.50	10.9
τ_{cool}	0.940	2.13	13.7	38.2	91.5	8.58×10^2	2.76×10^3
α_{cool}	-9.55	-7.50	-5.71	-5.19	-4.72	-0.359	1.02
$\mathcal{M}_{\text{cool}}$	7.03×10^9	1.07×10^{10}	3.72×10^{10}	1.10×10^{11}	4.24×10^{11}	1.23×10^{13}	2.85×10^{13}
$\Delta \log_{10} \mathcal{M}_{\text{cool}}$	0.307	0.547	2.07	2.77	3.39	6.19	14.6
β_{cool}	0.0505	0.669	3.80	5.34	6.46	7.78	8.24
V_{reion}	29.0	31.4	39.2	44.3	49.8	58.3	60.6
μ	-0.167	0.652×10^{-2}	0.186	0.253	0.327	0.467	0.510
κ	-0.313	-0.159	-0.896×10^{-2}	0.0401	0.0903	0.209	0.297
C_1	-2.69	-1.66	-0.467	0.0357	0.535	1.80	2.76
C_2	-2.88	-1.73	-0.458	0.0440	0.524	1.67	2.56
C_3	-2.79	-1.83	-0.495	-0.0652	0.362	1.54	2.60
C_4	-2.99	-1.83	-0.645	-0.186	0.274	1.50	2.51
C_5	-2.79	-1.73	-0.503	0.199×10^{-2}	0.478	1.64	2.60
C_6	-2.71	-1.69	-0.557	-0.0774	0.423	1.72	2.83
H_1	-2.86	-1.78	-0.559	-0.0416	0.441	1.69	3.04
H_2	-2.89	-1.77	-0.476	0.0156	0.463	1.59	2.66

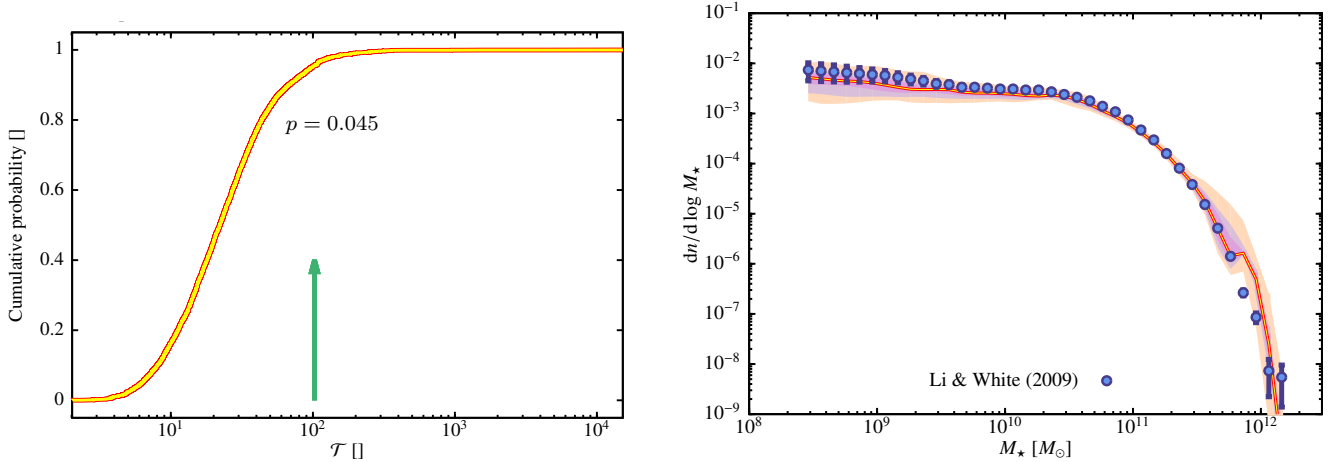


Figure 7. *Left panel:* The cumulative probability of the test statistic used in our PPC analysis for models sampled from the converged MCMC chains, \mathcal{T}_l (yellow line). The green arrow shows the same test statistic computed for the data, \mathcal{T}' . The resulting p -value (also indicated in the panel) is 0.045. *Right panel:* The stellar mass function of galaxies at $z \approx 0.07$. Blue points with error bars show the mass function measured from the SDSS by Li & White (2009). Error bars on these points are the square-roots of the diagonal elements of the covariance matrix constructed as described in §3.1. The shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3-“ σ ” respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models.

4.3 Effects of Uncertainties

To assess the importance of the various uncertainties that we account for in our analysis we perform additional MCMC simulations in which we successively remove each source of uncertainty. These simulations were performed in the same

manner as described in §3.7.3, except that 144 parallel chains per simulation were used and the chains were initialized from

the posterior of the previous simulation²⁵. Throughout this discussion we will refer to the marginalized posterior probability distributions shown in Fig. 8 which illustrate the range of responses of posterior probability distributions to these uncertainties.

4.3.1 Cosmological Parameters

We first perform an MCMC simulation in which the cosmological parameters $C_{1...6}$ are fixed to zero, such that all cosmological parameters are fixed at their maximum likelihood values from WMAP-9. We find that the posterior probability distributions of model parameters are not significantly affected by fixing cosmological parameters to the maximum likelihood values²⁶, as can be seen in all three panels of Fig. 8. Consequently, the model remains a viable description of the data even when cosmological parameters are held fixed.

4.3.2 Halo Mass Function Systematics

Keeping the cosmological parameters fixed, we next fix the halo mass function systematic parameters, $H_{1...2}$ to zero and repeat our MCMC analysis. The effect of these nuisance parameters on the posterior probability distributions is also quite small. There is a small change in the posterior probability distribution of the f_{wind} parameter (see top panel of Fig. 8) in which the probability is suppressed at high values of f_{wind} , and a similar effect on the posterior probability distribution of α_* (middle panel of Fig. 8). In general, we find that inclusion of halo mass function systematic uncertainties lead to small increases in the widths of posterior probability distributions but no significant bias.

4.3.3 Stellar Mass Systematics

We next remove stellar mass systematics by fixing μ and κ at the maximum likelihood values of their priors²⁷ and repeating our MCMC analysis. We find that the inclusion of these systematics can dramatically alter the posterior probability distribution for some parameters, while leaving others unaffected. For example, in Fig. 8 it is clear that the posterior probability distribution of f_{wind} is unaffected by stellar mass systematics, while the posterior probability distribution of α_* is strongly affected, with the peak of the posterior probability distribution shifting from 2 to 0.5. We note that, in our full analysis, the peaks of the posterior probability distributions for these parameters are located at $\mu \approx 0.22$ and $\kappa \approx 0.035$ respectively. These are both far from the mean of their priors, so it is not surprising to find a large shift in the posterior probability distributions of other parameters once μ and κ are no longer allowed to vary freely.

²⁵ Since in each successive simulation we are removing nuisance parameters this means that the chains still begin in an overdispersed state as required by the Gelman-Rubin convergence criterion.

²⁶ This does not imply that this would be the case for other constraints of course. In particular, it is likely that certain parameters (e.g. σ_8) have a much stronger effect at higher redshifts.

²⁷ Note that this means $\mu = -0.0125$ and $\kappa = 0.165$ since we retain the discrepancy term from §3.5.4.

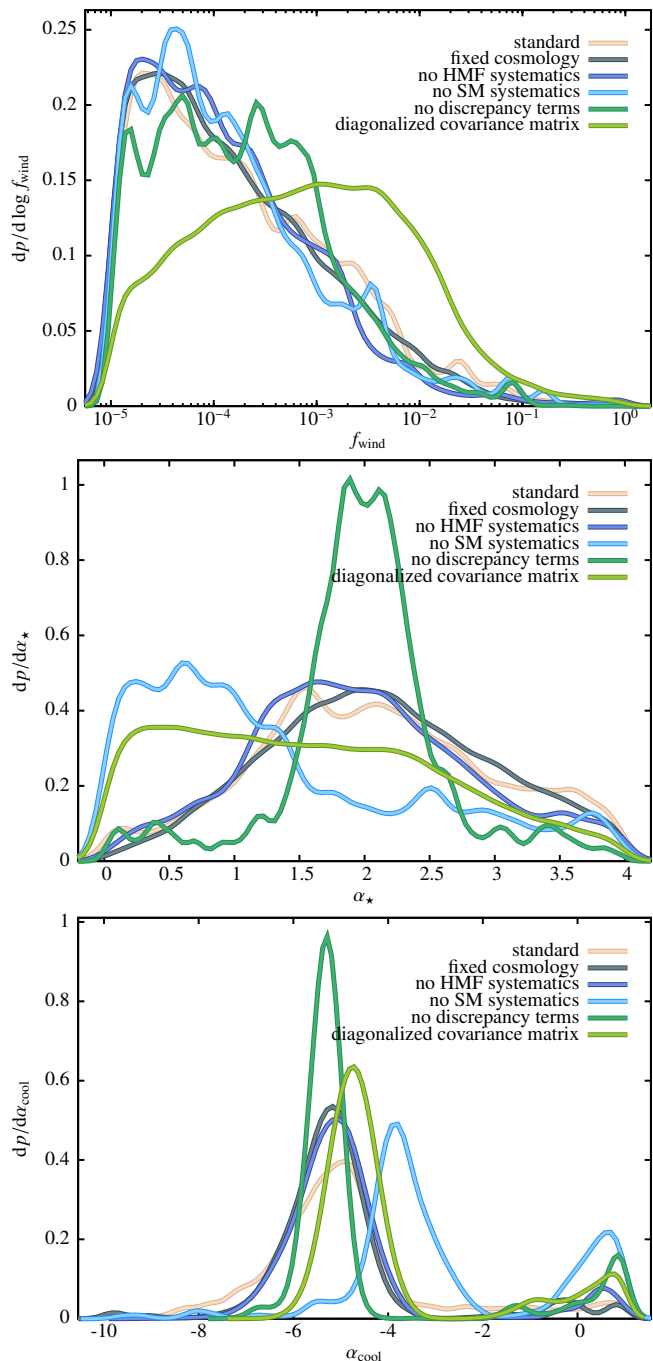


Figure 8. Marginalized posterior probability distributions on two parameters (left panel: f_{wind} ; right panel: α_*) from our MCMC chains. Line colors correspond to different modeling assumptions. From top to bottom: “standard” is our full calculation; “fixed cosmology” fixes cosmological parameters to the maximum likelihood values of their priors (see §4.3.1); “no HMF systematics” fixes halo mass function systematic parameters to the maximum likelihood values of their priors (see §4.3.2); “no SM systematics” fixes stellar mass systematic parameters to the maximum likelihood values of their priors (see §4.3.3); “no discrepancy terms” removes all model discrepancy terms (see §4.3.4); finally, “diagonalized covariance matrix” diagonalizes the covariance matrix before computing model likelihoods (see §4.3.5)

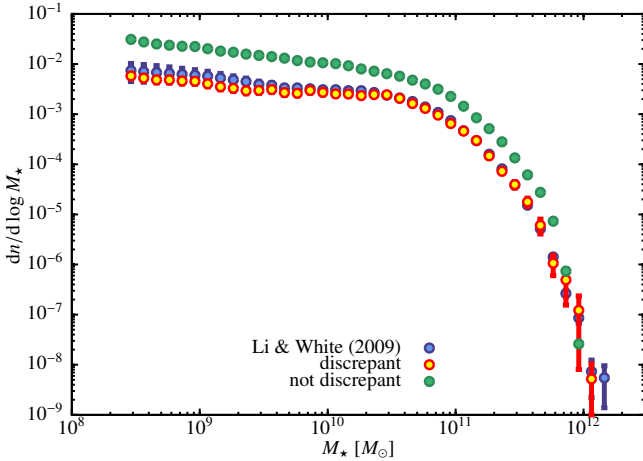


Figure 9. The maximum likelihood model evaluated including model discrepancy terms (yellow points), together with the same model with discrepancies removed (green points).

4.3.4 Model Discrepancies

Next, we set all model discrepancy terms to zero (such that the stellar mass systematic parameters are now fixed at $\mu = 0$, $\kappa = 0$) and repeat our MCMC analysis. We find that the peaks of the posterior probability distributions now return close to their locations for our full analysis. However, the posterior probability distributions of several model parameters are now *much* narrower than in our full analysis. This indicates the cumulative effect of removing all sources of model and data systematic uncertainties. For example, in the middle panel of Fig. 8 we see that the width of the posterior probability distribution for α_* is reduced by a factor of approximately 2 compared to the full analysis. This highlights the importance of the inclusion of nuisance parameters to model systematic uncertainties in deriving robust constraints on model parameters (and, therefore, for being able to make robust predictions).

It is also instructive to check whether our parameterizations of the model discrepancy terms remain valid for the maximum likelihood model. This is not guaranteed to be the case—the model discrepancy terms were computed using our *a priori* model which turns out to be significantly different from the final maximum likelihood model. Therefore, in Figure 9 we plot the stellar mass function data (blue points), our maximum likelihood model computed exactly as in our MCMC analysis and with model discrepancy correction terms applied (yellow points). Additionally, we re-run the maximum likelihood model, but now remove all model discrepancies (i.e. we use distributions of virial orbits and merging times, and use a fixed merger tree resolution—note that we do not remove the discrepancy relating to the use of Parkinson et al. (2008) merger trees rather than N-body-derived merger trees, but this causes no offset in the model mass function), and do not include model discrepancy corrections when evaluating the mass function. The result is shown by the green points, which are clearly significantly offset from the model including discrepancies. This clearly shows that our parameterizations of model discrepancies—calibrated using our *a priori* model—do not well describe the model discrepancies in our maximum likelihood model.

In future work, we can use the maximum likelihood model from this work to assess model discrepancies for further MCMC analysis. This should result in more accurate parameterization of model discrepancies. However, the fundamental problem illustrated by Figure 9 is not resolved by this fact—most likely we will be evaluating model discrepancies for some other observable for which the present maximum likelihood model may not be the best description.

This suggests, for future studies, an iterative procedure. That is, model discrepancies are determined using the current maximum likelihood model. These are then used in a quick search of parameter space to find a new maximum likelihood model. If this is sufficiently different from the previous maximum likelihood model (where “sufficiently different” here means that the model discrepancy terms differ substantially), then it is used to evaluate new model discrepancies. This process is repeated until convergence on the model discrepancies is reached.

4.3.5 Diagonalized Covariance

Finally, we explore the importance of off-diagonal elements of the covariance matrices by simply setting all off-diagonal elements to zero and repeating our MCMC analysis²⁸. This throws away any information about correlations between bins in the stellar mass function (both observed and modeled). As we have shown in Fig. 3, these correlations are significant, and so we may expect diagonalizing the covariance matrix to significantly affect our results. This is in fact the case, as can be seen in all three panels of Fig. 8. Comparing the “no discrepancy terms” and “diagonalized covariance matrix” terms we see that the posterior is substantially shifted in all three cases when correlations between bins in the stellar mass function are ignored. In particular, for the parameter α_{cool} (bottom panel of Fig. 8) the shift is, approximating the posterior probability distribution as Gaussian, slightly over 1σ .

4.4 Predictions

Using the posterior probability distribution of our model, we can explore predictions for other observables. Given the phenomenological nature of the model described in §2 we in fact should consider these to be “extrapolations” rather than “predictions”. True “predictions” will only arise from a model with a more physical basis.

When comparing the PPC of the model posterior with these datasets we must be careful to account for the errors in the observed data. This is not a problem when carrying out PPC on the dataset used to constrain the model because the model posterior implicitly accounts for the observational errors on that dataset. When comparing to other datasets though it is possible that those datasets may have much larger errors. Computing the test statistic at only the measured values of the data would then likely lead to an extreme value for the test statistic, suggesting inconsistency with the model.

²⁸ A similar study of the importance of correlated errors was carried out by Lu et al. (2011), although using a phenomenological model of the covariance matrix.

To avoid this problem, we perturb each model realization by a random realization of the errors in the observed data, using the covariance matrix (assumed to be diagonal) of the data to perturb around the true values. The distribution of the test statistic, and the associated p -value are then computed in the usual way.

Figure 10 compares our model to the galaxy HI mass function measured from the ALFALFA survey (Martin et al. 2010), while Figure 11 compares our model to stellar mass functions out to $z \approx 1$ from the PRIMUS survey (Moustakas et al. 2013).

5 DISCUSSION

Bearing in mind the phenomenological nature of our current model, it would be unreasonable to draw too much inference about the nature of galaxy formation physics from the posterior probability distribution of Figure 5. Nevertheless, there are a few interesting facts to be gleaned from the posterior probability distribution. To elucidate these points we apply the “projection pursuit” methodology of Bower et al. (2010) to identify the linear combinations of model parameters which contribute *least* to the variance of the posterior probability distribution. We follow Bower et al. (2010) in scaling each parameter by the width of its prior when carrying out the PCA analysis. Specifically, for uniform priors we scale parameters by the difference between the upper and lower limits of the prior, while for normal priors we scale by the root variance. Table 4 shows the results of this analysis.

From the posterior probability distribution and projection pursuit analysis we find:

- The exponent of the virial velocity dependence of the star formation efficiency, β_* , is tightly constrained to the range $[-4.05, -3.24]$ (68.3% confidence interval) with maximum likelihood value -3.59 . As can be seen in Table 4, β_* is the dominant contributor to the second principal projection (which has an additional contribution from β_{cool}). We explore what drives this tight constraint on this parameter combination by perturbing the maximum likelihood model along the direction of this principal projection²⁹ in parameter space by ± 1 , 2, and $3\sqrt{V}$ where V is the variance of the principal projection³⁰. The results of this study are shown in the right-hand panel of Figure 12, in which we plot the model stellar mass function relative to the observed mass function and normalized by the root-variance of the observed mass function. This indicates that it is the “knee” and high-mass cut-off of the mass function which is strongly affected by this parameter combination (and, therefore, which is driving the tight constraint on these parameters). Note in particular that the perturbation changes sign at around $2 \times 10^{11} M_\odot$ —the constraint is driven primarily by the shape of the knee/cut-off region of the mass function.

- The very flat slope of the stellar mass function at low masses leads to a preference for very negative exponents

in the virial velocity dependence of the outflow rate (see eqn. 7), β_{wind} . Values as low as $\beta_{\text{wind}} = -14.95$ are preferred (even more negative values are allowed, and are limited mostly by the imposed prior), although values $\beta_{\text{wind}} < -0.57$ are not ruled out at greater than 99.7% confidence. This illustrates the difficulty of obtaining a sufficiently flat galaxy mass function given the very steep slope of the halo mass function.

Previous SAMs have also found the need for a strong dependence of outflow mass loading on halo virial velocity. For example, Cole et al. (1994) and Bower et al. (2010) required values $\beta_{\text{wind}} = -5.5$ and -3.2 respectively, while Lu et al. (2013) find $\beta_{\text{wind}} \approx 6$ is required to produce slopes shallow enough to match their constraints. There is a correlation between β_{wind} the normalization of the outflow rate, f_{wind} , and the redshift exponent of outflow, α_{wind} , (projection 9 in Table 4) such that less negative values of β_{wind} require larger values of the normalization. This is fortunate as physical models of outflows typically predict values β_{wind} in the range -2.7 to -1 (Murray et al. 2005; Oppenheimer et al. 2010; Creasey et al. 2013; Lagos et al. 2013), while observations indicate mass loading factors for galaxies of stellar mass $\sim 10^{10} M_\odot$ of order unity (Martin et al. 2012). In this current, sufficiently simple, model, outflows can never actually remove gas from a halo (instead simply returning it to the hot component within the halo). An extended model in which outflows can actually leave the halo (returning only if the halo later grows substantially in mass) might reduce or remove the need for such a strong scaling with halo virial velocity.

- The exponent of the cooling timescale, α_{cool} , is tightly constrained to lie in the range -6.12 to -4.35 at 68.3% confidence with maximum likelihood value -5.12 . As shown in Table 4, α_{cool} is the dominant contributor to the projection with least variance, with other significant contributions from the cooling timescale, τ_{cool} , and the exponent of the cut-off in cooling efficiency, β_{cool} . In the left-hand panel of Figure 12 we show the effect of perturbing the maximum likelihood model along this parameter projection. Not surprisingly, these perturbations affect the high mass end of the mass function, shifting it systematically above or below the observed mass function. This parameter combination directly affects the masses of high-mass galaxies.

- Our model includes a parameterization of a cut-off in cooling efficiency in high mass halos. We find that a relatively low mass cut-off is preferred (maximum likelihood value of $10^{10.1} M_\odot$, with 68.3% of the posterior lying between $10^{10.0}$ and $10^{11.6} M_\odot$), although the posterior extends to cut-off masses as high as $10^{13.4} M_\odot$ at 99.7% confidence. This cut-off has often been associated with AGN feedback, or the transition from “cold-mode” to “hot-mode” cooling. Our posterior on the cut-off mass is consistent with the expected mass scales for these processes (Birnbom & Dekel 2003; Kereš et al. 2005), but is too broad to place a meaningful constraint.

- The combination of the width of the cut-off in cooling efficiency, $\Delta \log_{10} \mathcal{M}_{\text{cool}}$, and the exponent of the cut-off, β_{cool} , is, however, very well constrained. Cooling must transition from being efficient to inefficient over about 1 order of magnitude in halo mass if β_{cool} is around unity. Broader transitions are allowed if the cut-off is made sharper by rais-

²⁹ When performing this perturbation we include contributions from all parameters to this principal projection, not just the dominant contributors listed in Table 4.

³⁰ Of course, we remove the scaling of the parameters used in the PCA analysis when constructing the perturbation.

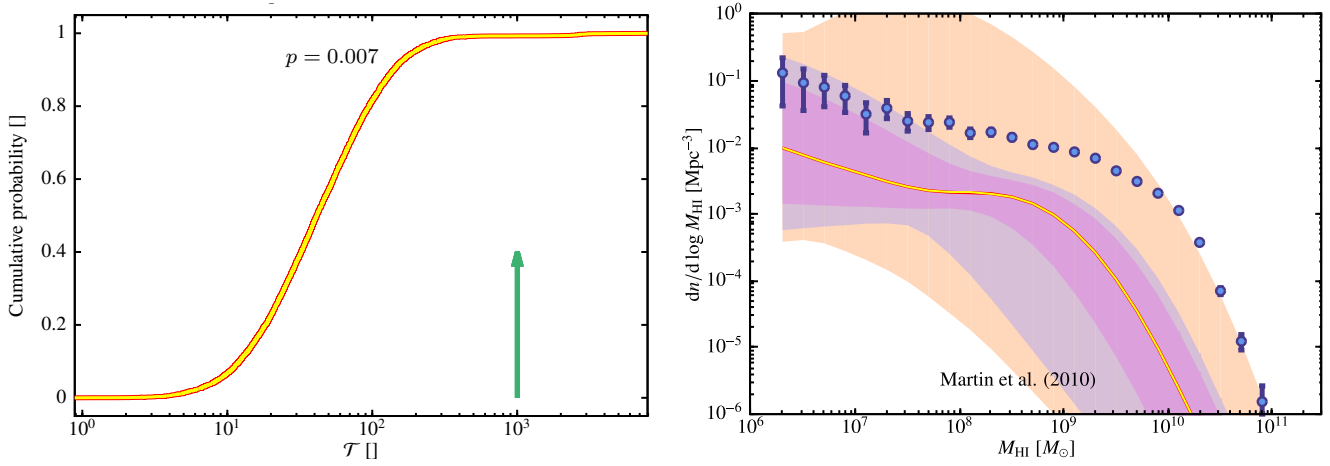


Figure 10. *Left panel:* The cumulative probability distribution of the test statistic used in our PPC analysis for the HI galaxy mass function at $z = 0$ using models sampled from the converged MCMC chains, \mathcal{T}_l (yellow line). The green arrow shows the same test statistic computed for the data, \mathcal{T}' . The resulting p -value (also shown in the panel) is 0.007. *Right panel:* The $z = 0$ HI galaxy mass function of Martin et al. (2010) compared to the extrapolation from our model. Shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3- σ respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models.

Table 4. Results of a PCA analysis of the model posterior using the “projection pursuit” methodology of Bower et al. (2010). Rows list principal components, arranged in order of least to most contribution to the variance in the posterior (as listed in the second column). For each principal component, model parameters which contribute significantly to the variance of the component (greater than 10%) are listed, along with the fraction of the variance in the component that they contribute.

Projection	Variance	Contributions		
0	0.13×10^{-2}	$0.56 \alpha_{\text{cool}};$	$0.26 \tau_{\text{cool}};$	
1	0.16×10^{-2}	$0.73 \Delta \log_{10} \mathcal{M}_{\text{cool}};$		
2	0.26×10^{-2}	$0.39 \beta_{\text{wind}};$	$0.30 \beta_{\star};$	
3	0.69×10^{-2}	$0.57 \beta_{\star};$		
4	0.95×10^{-2}	$0.56 \tau_{\star};$	$0.21 \alpha_{\star};$	$0.10 \beta_{\star};$
5	0.21×10^{-1}	$0.30 \tau_{\text{cool}};$		
6	0.38×10^{-1}	$0.41 f_{\text{wind}};$	$0.15 \beta_{\text{wind}};$	
7	0.41×10^{-1}	$0.78 \tau_{\star, \text{min}};$	$0.11 f_{\text{wind}};$	
8	0.59×10^{-1}	$0.39 \alpha_{\star};$		
9	0.72×10^{-1}	$0.33 V_{\text{reion}};$	$0.12 \alpha_{\text{wind}};$	
10	0.76×10^{-1}	$0.44 V_{\text{reion}};$	$0.31 \tau_{\text{wind, min}};$	
11	0.76×10^{-1}	$0.33 f_{\text{df}};$	$0.15 \mathcal{M}_{\text{cool}};$	$0.12 \tau_{\text{cool}}; \quad 0.11 \alpha_{\star};$
12	0.86×10^{-1}	$0.43 \tau_{\text{wind, min}};$	$0.38 \mathcal{M}_{\text{cool}};$	
13	0.90×10^{-1}	$0.49 \alpha_{\text{wind}};$	$0.20 f_{\text{df}};$	$0.15 \beta_{\text{wind}};$
14	0.10	$0.46 \beta_{\text{cool}};$		
15	0.20	$0.87 \mu;$		
16	0.81	$0.85 \kappa;$		
17	0.84	$0.35 C_3;$	$0.21 C_6;$	$0.17 C_2;$
18	0.89	$0.40 C_1;$	$0.39 H_2;$	
19	0.95	$0.23 H_2;$	$0.23 C_3;$	$0.12 H_1;$
20	0.96	$0.41 C_4;$	$0.33 H_1;$	
21	0.98	$0.81 C_5;$	$0.11 H_1;$	
22	1.1	$0.13 C_1;$	$0.11 H_1;$	
23	1.1	$0.35 C_2;$	$0.23 H_1;$	
24	1.2	$0.27 C_1;$		

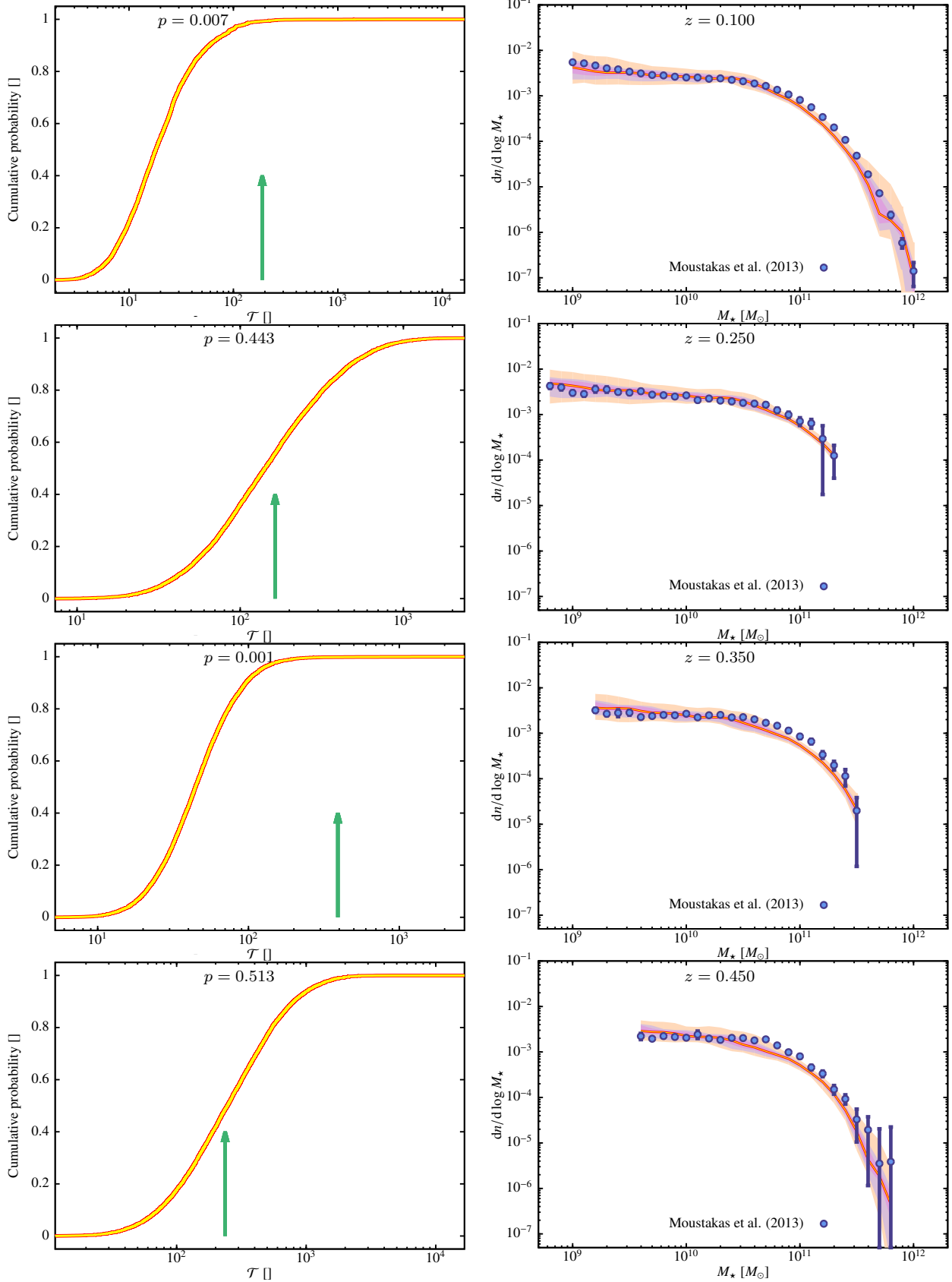


Figure 11. *Left panel:* The cumulative probability distribution of the test statistic used in our PPC analysis for models sampled from the converged MCMC chains, \mathcal{T}_i (yellow line), for stellar mass functions measured from the PRIMUS survey (Moustakas et al. 2013) at the redshifts indicated in each panel. Where visible, the green arrow shows the same test statistic computed for the data, \mathcal{T}' . The resulting p -value is 0.007, 0.443, 0.001, 0.513. *Right panel:* Stellar mass functions from the PRIMUS survey at redshifts as indicated in the panels (blue points). The shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3-“ σ ” respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models.

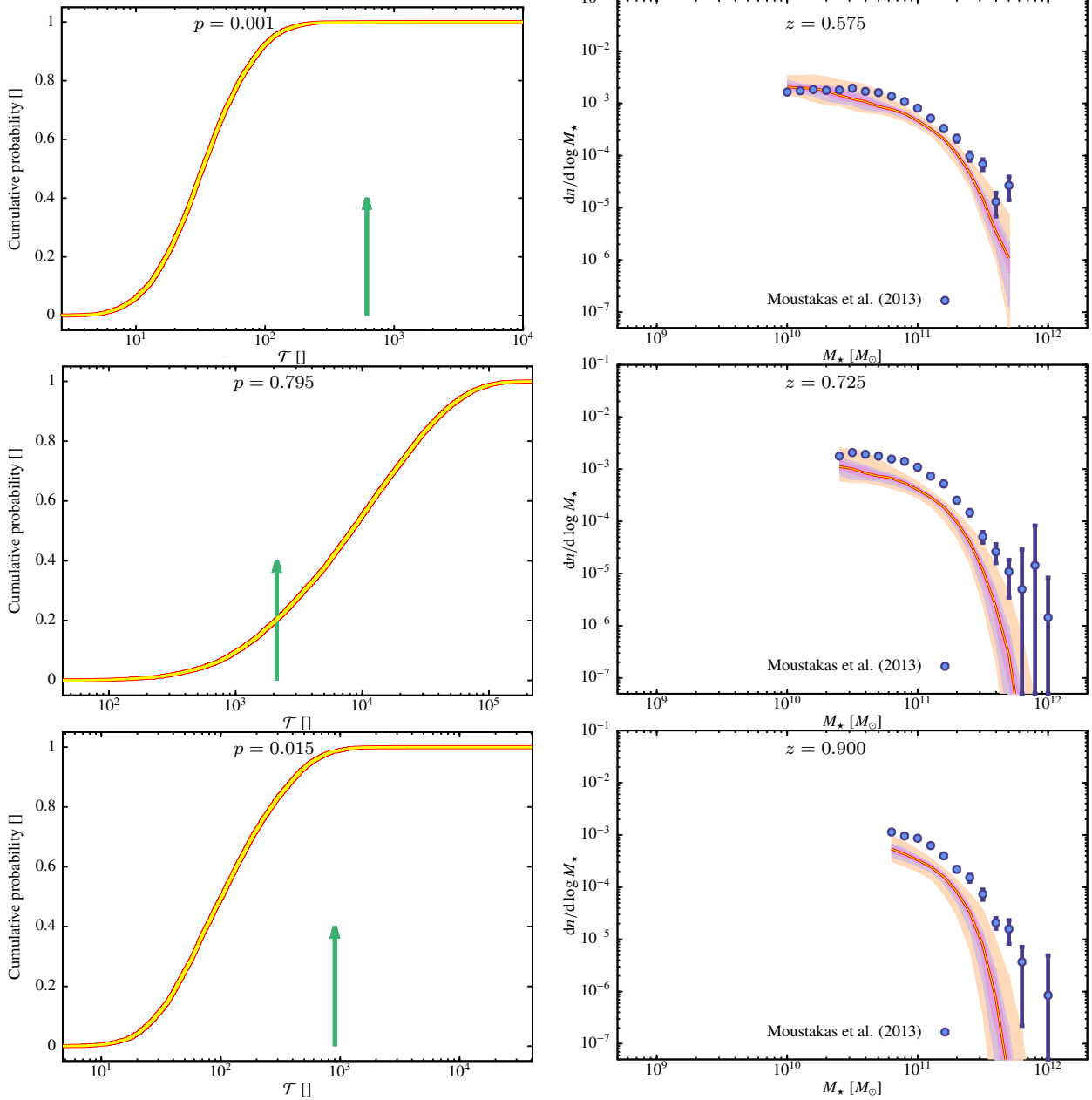


Figure 11. (*cont.*) p -values for these redshifts are 0.001, 0.795, and 0.015.

ing the exponent, β_{cool} —a result of the strong correlation between $\Delta \log_{10} \mathcal{M}_{\text{cool}}$ and β_{cool} .

As mentioned in §4.2 the median model mass function drawn from the posterior probability distribution lies systematically below the data at the low-mass end of the stellar mass function. This is a consequence of two facts. First, the data show a weak, but significant change in slope at around $M_* \approx 4 \times 10^9 M_\odot$. Our model contains nothing which would allow it to reproduce such a feature (i.e. in this regime the model is essentially self-similar). The second fact is that the covariance matrix of the data has strong off-diagonal correlation at low-masses (a consequence of large scale structure contributions to the covariance). Combined, these facts mean that the model can achieve a high likelihood with a constant slope at low masses because the low-mass data

points effectively move up and down in unison, rather than as independent points. Colloquially speaking, the model lies roughly “1- σ ” below each of the low mass points, but this corresponds to a likelihood consistent with being “1- σ ” below a single point. Figure 13 shows the model stellar mass function computed when the covariance matrix used in our likelihood calculation is diagonalized. In this case the model is forced to go through all points in the low mass part of the mass function, effectively overfitting the data. This illustrates the importance of constructing the model likelihood using the full covariance matrix of the data, rather than just a diagonalized approximation to that matrix. Similar effects have been seen by Lu et al. (2013) when constraining their model to fit the HI mass function of galaxies.

When our model is extrapolated to higher redshifts we find that, while it is formally not a good description of the

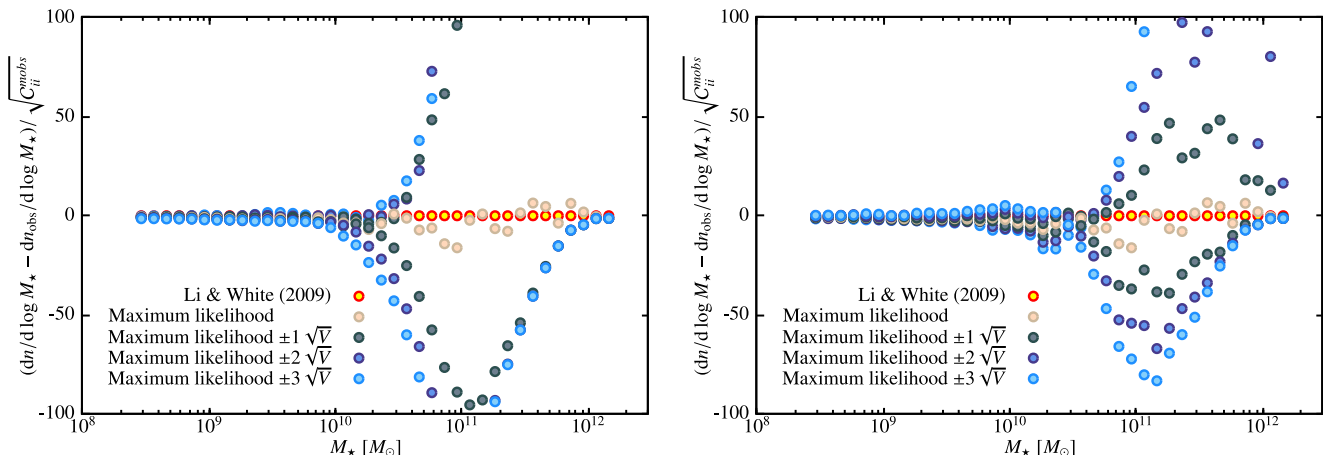


Figure 12. Models perturbed away from the maximum likelihood model along the directions of projections 0 and 2 (from Table 4) by 0, ± 1 , ± 2 , and ± 3 times the root variance of the projection. The points show the offset of the model mass function from the observed mass function normalized by the root-variance of each of the observed mass function.

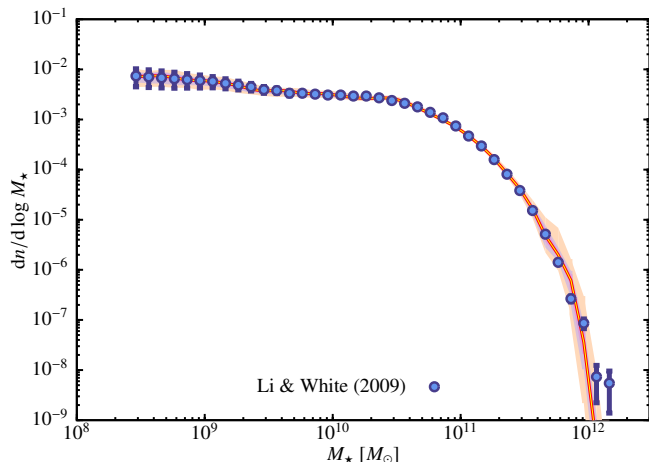


Figure 13. The stellar mass function of galaxies at $z \approx 0.07$ when the model is constrained using a diagonalized covariance matrix. Blue points with error bars show the mass function measured from the SDSS by Li & White (2009). Error bars on these points are the square-roots of the diagonal elements of the covariance matrix constructed as described in §3.1. The shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3- σ respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models.

stellar mass function out to $z \approx 1$, given that we have not yet performed a detailed analysis of random and systematic errors in those mass functions (as we have done for $z \approx 0$), the model is remarkably successful in capturing the evolution of the shape and normalization of the stellar mass function, even though it is constrained only by $z \approx 0$ data³¹. For ex-

ample, at $z = 0.9$ the 99.7% confidence interval of the model comes close to several of the data points. We can compare this to the model extrapolation if we, for example, keep cosmological parameters fixed at their maximum likelihood values rather than letting them vary with the WMAP 9-year priors. The left panel of Fig. 14 shows the $z = 0.9$ stellar mass function in this case. The range spanned by the model extrapolations is clearly much narrower than that shown in Figure 11 (in which cosmological parameters were allowed to vary). Particularly at high redshifts, the remaining uncertainties in cosmological parameters contribute significantly to the uncertainty in theoretical models constrained by $z = 0$ data. Even bigger changes are apparent if we set all systematic uncertainty parameters to zero and ignore model discrepancies, as shown in the right panel of Fig. 14. In this case the range of model predictions at $z = 0.9$ is dramatically reduced (although no significant bias is introduced), clearly indicating the importance of systematic uncertainty analysis when extrapolating models constrained by one dataset to make predictions for another.

In comparing our model with the HI galaxy mass function from the ALFALFA survey (Martin et al. 2010) it is clear that the model predictions are both highly unconstrained³², and, in most instances, dramatically underpredict the mass function. Our test statistic indicates an extremely poor agreement between model and data, as expected. This disagreement between model and HI mass function suggests that adding the HI mass function as a constraint on our model should provide significant additional constraining power, although additional freedom may have to be introduced into the model to allow a good fit to be found. We again caution that these results should not be considered final as we have not yet accounted for system-

³¹ After the MCMC calculations for this work were underway, Bernardi et al. (2013) published a new measurement of the SDSS galaxy stellar mass function, based on a new analysis of the photometry. They conclude that the galaxy magnitudes used by Li & White (2009) underestimate the total luminosities of the brightest galaxies. As such, Bernardi et al. (2013) find that the high-mass end of the mass function should be boosted relative to Li & White (2009). If our model were fit to the result of Bernardi et al. (2013)

instead of to those of Li & White (2009) we might therefore expect improved agreement between our model and $z \approx 1$ stellar mass functions. This will be explored in the next paper in this series.

³² This indicates that, within the structure of our model, it is possible to obtain the correct stellar mass function from galaxy populations with vastly different gas content.

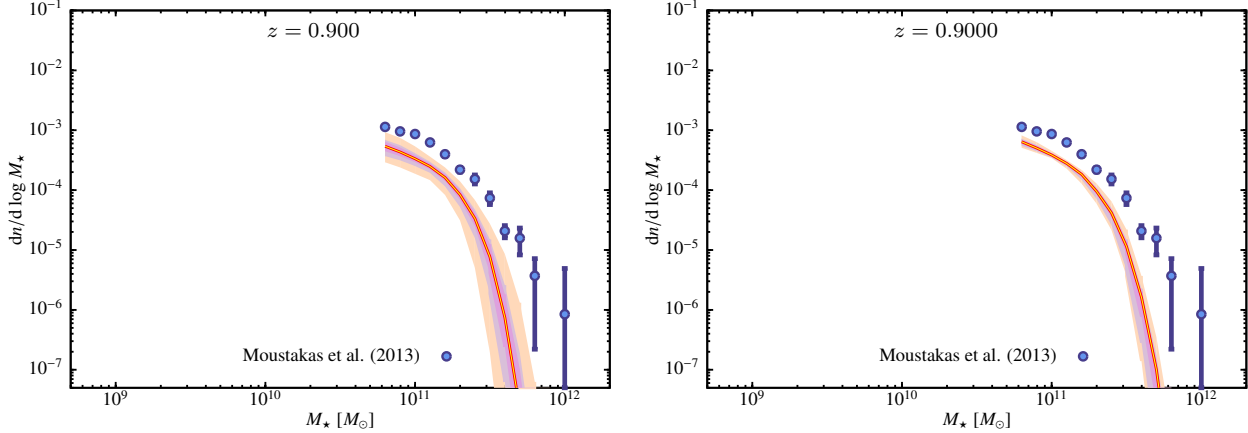


Figure 14. Stellar mass function from the PRIMUS survey at $z = 0.9$ (blue points). The shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3- σ respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models. *Left panel:* Models in which cosmological parameters were held fixed at their maximum likelihood values. *Right panel:* Models in which all systematic uncertainties and model discrepancy terms were set to zero.

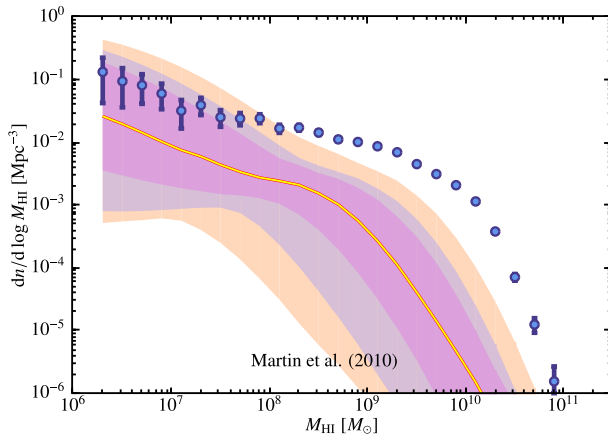


Figure 15. The $z = 0$ HI galaxy mass function of Martin et al. (2010) compared to the extrapolation from our model when all systematic uncertainty parameters are set to zero and model discrepancy terms are ignored. Shaded regions enclose the inner 68.26%, 95.44%, and 99.74% (1, 2, and 3- σ respectively) of the distribution of models sampled from the converged MCMC chains, while the yellow line indicates the median of those models.

atics in the determination of HI masses, nor have we assessed how model discrepancy terms would impact the HI mass function. In particular, the very coarse resolution of our merger trees might be expected to substantially affect the predicted HI mass function. These issues will be explored in the next paper in this series. Once again, the inclusion of systematic uncertainties in our modelling strongly affects the distributions of model HI mass functions, as can be seen in Fig. 15 in which we show the distribution of model HI mass functions when systematics and model discrepancies are ignored (compare with Fig. 10).

5.1 Comparison with previous works

Several previous works have applied MCMC or similar techniques to SAMs with the goal of constraining the parameters of these models. In the following we discuss several of these

works and highlight the similarities and differences with the current work.

Henriques et al. (2009) explored a 6-dimensional model parameter space using MCMC techniques. They find that they are able to obtain reasonable fits to a combination of K-band luminosity function, B–V colours, and the black hole mass function all at $z \approx 0$, but find that this requires very strong feedback from supernovae, as do we in the present work. Importantly, Henriques et al. (2009) highlight a discrepancy between their model constrained to the K-band luminosity function and the stellar mass function, which they take to suggest that a possible 0.3 dex systematic in stellar mass determinations may exist. Such systematics are modeled directly in our calculation. A systematic of 0.3 dex is entirely consistent with our derived posterior distributions for the parameters μ and κ of our stellar mass systematics model. Importantly, the constraints on all of our other parameters correctly take into account the full range of possible systematics. Henriques et al. (2009) also acknowledge the importance of systematics in assessing the validity of their model—in the current work we are able to address this issue directly. Continuations of this work (Guo et al. 2011; Henriques et al. 2013) explores constraints derived from luminosity and stellar mass functions over the redshift range $0 < z < 3$. Henriques et al. (2013) employ an 11-parameter model and find that the efficiency of reaccretion of ejected material must explicitly vary with cosmic epoch in order to fit the constraints. Once constrained in this way, their model reproduces quite well the evolution of galaxy abundances, the colours of $z = 0$ dwarf galaxies, and the clustering of galaxies across a wide range of masses. Following Henriques et al. (2009), Henriques et al. (2013) attempt to address the presence of systematic errors in datasets by using the scatter between multiple observational determinations of a particular constraint. This approach has the advantage that it does not require assumptions about the sources of systematics to be made. However, as noted by Henriques et al. (2013) it does require arbitrary judgements about the quality of different datasets to be made and, furthermore, will not account for systematics in common between datasets. As a result, Henriques et al. (2013) conclude that “*As a result, formal*

levels of agreement between theory and observation should be treated with caution.” Clearly, consideration of how to treat systematic uncertainties is crucial for progress in this area.

Bower et al. (2010) explored the 16-dimensional parameter space of their model using a model emulator technique, constraining their model to fit $z \approx 0$ bJ and K-band luminosity functions. They found, as do we, that a wide range³³ of parameter space is compatible with the observational data. Bower et al. (2010) take into account three model discrepancy terms. The first arises from their choice to exclude eight “inactive parameters” (parameters to which the results of the model are only weakly dependent) from their model emulator—this discrepancy accounts for the additional variance that those parameters would introduce were they included in the emulation. Since we treat all parameters as active in the present work we have no comparable discrepancy. The second discrepancy included by Bower et al. (2010) accounts for the finite volume of the simulation volume from which they extract their merger trees. This is accounted for in the present work when modeling the differences between Monte Carlo and N-body merger trees³⁴; see §3.5.3. Finally, Bower et al. (2010) include a discrepancy term to account for the structural deficiencies of their model. This is comparable to the discrepancies that we introduce in §3.5. Bower et al. (2010) adopt a rather different approach to estimating this discrepancy, namely exploring a range of model parameter space and judging which models they would consider to be adequate fits to the data based on previous experience. While this approach has its advantages compared to our own (namely being applicable even in cases where the “correct” answer cannot be modeled), it is difficult to apply to constraints for which previous intuition is lacking.

Mutch et al. (2013) constrain a 6-dimensional SAM to simultaneously fit $z \approx 0$ and $z \approx 0.83$ galaxy stellar mass functions. They also find that extremely strong supernova feedback is required to match these datasets, and note a minor tension between the model and data when fitting both redshifts together. Mutch et al. (2013) suggest that this tension may be removed by more accurate estimation of random and systematic errors in the data. In this work, we have shown that these errors do in fact make a significant difference to the constraints on models, and may help reconcile the tension found by Mutch et al. (2013). We note, however, that our own models remain formally inconsistent with the high redshift stellar mass functions explored here, with the caveat that we have so far applied a rigorous error analysis only to the $z \approx 0$ stellar mass function. Once a similar analysis is applied to the higher redshift data it remains possible that our model will be consistent with all data considered in this work.

Finally, Lu et al. (2013) constrain a 17-dimensional SAM to fit a K-band luminosity function and HI galaxy mass function. Once again, a requirement for very strong supernova driven feedback is found, and that a very strong dependence on halo mass (or virial velocity) is required to

produce a sufficiently shallow faint end slope (consistent with our own findings regarding the slope of the stellar mass function at low masses). Lu et al. (2013) employ PPCs to assess whether their model is a viable description of the data, finding that it is not (the data lies well into the tails of the distribution of model results), similar to our own findings. This suggests that the detailed form of these particular datasets requires more refined models (possibly containing additional physics).

To summarize, a consistent picture is beginning to emerge from studies of this sort, with conclusions including the need for very strong supernova feedback, and the inability of current models to describe existing data with the accuracy required by observational uncertainties. In this work we have extended previous studies by taking into account the full range of model and data errors—both random (including covariances) and systematic—and sources of model discrepancy. While this results in a very high (25) dimensional model and requires additional modeling to assess observational and theoretical errors we have demonstrated that such an approach is feasible given current computational abilities, and is important if predictions from such models are to be used to robustly test galaxy formation theory.

6 CONCLUSIONS

We have performed a careful study of how to constrain a semi-analytic model of galaxy formation to a single observational dataset—specifically the stellar mass function of galaxies in the nearby Universe. We have paid particular attention to assessing the covariance matrices of data and model, both random and systematic errors in the data, and discrepancies in the model itself. Combining these multiple sources of uncertainty we derive a robust posterior probability distribution on the model parameters.

We find that, of these various sources of uncertainty, cosmological parameters and halo mass function systematics are sufficiently well constrained by other experiments to not significantly affect the specific observations to which we constrain our model in this work. However, the remaining uncertainties in cosmological parameters *do* play a crucial role in setting the uncertainty of theoretical model expectations for predictions at higher redshifts. Systematic uncertainties in stellar mass estimates are the largest source of uncertainty in our model expectations for higher redshifts. Understanding these systematic errors in stellar mass, and how they vary with redshift, will be crucial for robustly constraining and testing models in future. Ignoring this contribution to the model uncertainty could result in viable models being erroneously ruled out³⁵. We also find that ignoring correlations between bins in the stellar mass function when

³³ Although still small compared to the prior range of parameter space.

³⁴ Additionally, in this work, we account for the finite number of merger trees used in our model evaluation by directly modeling the model covariance matrix; see Appendix D.

³⁵ Of course, overestimating uncertainties can result in the opposite problem, namely forgiving models that are inadequate. This should be avoided to, but we believe that erring on the size of larger uncertainties is the more conservative way to proceed. Eventually, we would like to use these model frameworks to test specific hypotheses relating to aspects of galaxy formation physics. We would prefer to wrongly accept a bad model (which can always be ruled out later with the addition of better data)

computing model likelihoods (i.e. using only the diagonal terms in the covariance matrix) can substantially bias the resulting posterior probability distributions.

Given the phenomenological nature of our current model we refrain from drawing strong conclusions about the physics of galaxy formation. However, we find that the very flat slope of the galaxy stellar mass function is particularly challenging to reproduce within the framework of our model, requiring a very strong dependence of outflow efficiency on halo virial velocity. Furthermore, we find that the cooling of gas in halos must transition from efficient to inefficient over around one order of magnitude in halo mass (of less, depending on the functional form of the transition).

We compare our model, constrained to match the stellar mass function of galaxies in the local Universe, to other datasets. We find that it succeeds in matching the evolution of the galaxy stellar mass function out to $z \approx 1$ —a remarkable fact given that the model was not constrained by any high-redshift data. This suggests that the evolution of the stellar mass function is encoded in the present day mass function plus basic cosmological structure formation³⁶. However, we find that our model is not in agreement with the galaxy HI mass function in the local Universe, dramatically under-predicting the HI masses of galaxies. We note that Lu et al. (2013) were also unable to find a model which came close to the observed HI mass function, although in their case the model dramatically over-predicted the observations. We caution that this failure should not be taken too seriously as yet—we have not yet accounted for systematic uncertainties in HI mass determinations, nor for the limitations of the model in predicting HI mass. Furthermore, it is possible that a mode exists in the model parameter space which adequately describes both the stellar and HI mass functions, yet has too low likelihood to be populated when the stellar mass function alone is used as a constraint (Lu et al. 2012, 2013). Nevertheless, this failing will be our focus in a future work.

Both Mutch et al. (2013) and Lu et al. (2013) find significant tensions between their SAMs and observational data (the evolution of the luminosity function, and $z = 0$ luminosity function plus HI mass function respectively). These results illustrate the power of MCMC methods to uncover failings of models, as noted in particular by Lu et al. (2013). These conclusions are clearly of great importance for advancing our understanding of galaxy formation. Given this importance, we believe that it is crucial to carefully determine the systematic and random uncertainties of both models and data, and to account for these uncertainties when assessing likelihoods to provide accurate and robust constraints on model parameters. We will therefore revisit these additional constraints in the next paper in this series, including a careful analysis of random and systematic errors in those datasets.

Our current model relies almost entirely on phenomenological treatments of baryonic physics. This limits the insight that can be obtained from such models. Our goal is to progress swiftly to incorporating detailed physical mod-

els. To do this, it is crucial that any physical model (e.g. of star formation) be embedded within a model which gives a broadly correct description of the global properties of galaxies, such as their masses, sizes, densities, and dark matter halos. Otherwise, it will be difficult, if not impossible, to judge whether the failing of a model is due to an inadequate physical model, or simply due to that model being embedded within a galaxy population which looks nothing like that which is actually observed.

As such, the goal of the next paper in this series will be to construct a model which gets the basic structural properties (masses, sizes, densities, and halo masses) of galaxies correct in the local Universe and, where possible, out to moderate redshifts ($z \approx 1$). This will provide a sound basis on which to develop physical prescriptions. With such a model in hand it will become possible to augment the physics of the model and make quantitative inference about the validity of our understanding of galaxy formation physics.

Computational speed remains a serious issue for this type of study, in which tens of millions of model evaluations are required. Therefore, in the next paper in this series, we will also explore methods to mitigate this computational demand as will be required to make further studies of more realistic models tractable.

ACKNOWLEDGMENTS

We thank Richard Bower, Shaun Cole, Darren Croton, John Helly, Neal Katz, Yu Lu, Martin Weinberg, and Martin White for invaluable conversations. We would also like to thank Martin Weinberg for developing, sharing and providing assistance with the BIE package, Alexander Knebe for providing data on halo mass functions, and the VIRGO Consortium for making available data from the “MillGas” simulation. We acknowledge the hospitality of the International Centre for Radio Astronomy Research at the University of Western Australia, the Aspen Center for Physics (NSF grant #1066293), and the Institute for Computational Cosmology, at the University of Durham where parts of this work were completed.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

then to wrongly exclude a correct model which may then be no longer considered even as better data becomes available.

³⁶ At least, that is, given the particular structure of our model.

REFERENCES

- Adelman-McCarthy J. K., et al., 2008, *ApJS*, 175, 297
- Anderson L., et al., 2012, *MNRAS*, 427, 3435
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
- Benson A. J., 2005, *MNRAS*, 358, 551
- Benson A. J., 2010, *Physics Reports*, 495, 33
- Benson A. J., 2012, *NewA*, 17, 175
- Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38
- Benson A. J., Borgani S., De Lucia G., Boylan-Kolchin M., Monaco P., 2012, *MNRAS*, 419, 3590
- Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, *MNRAS*, 436, 697
- Beutler F., et al., 2011, *MNRAS*, 416, 3017
- Birnboim Y., Dekel A., 2003, *MNRAS*, 345, 349
- Blake C., et al., 2012, *MNRAS*, 425, 405
- Blanton M. R., et al., 2005, *AJ*, 129, 2562
- Bower R. G., Vernon I., Goldstein M., Benson A. J., Lacey C. G., Baugh C. M., Cole S., Frenk C. S., 2010, *MNRAS*, 407, 2017
- Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, *MNRAS*, 351, 1151
- Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, *MNRAS*, 271, 781
- Creasey P., Theuns T., Bower R. G., 2013, *MNRAS*, 429, 1922
- Croton D. J., et al., 2006, *MNRAS*, 365, 11
- Das S., et al., 2011, *ApJ*, 729, 62
- De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
- Font A. S., et al., 2011, *MNRAS*, 417, 1260
- Freedman W. L., Madore B. F., Scowcroft V., Burns C., Monson A., Persson S. E., Seibert M., Rigby J., 2012, *ApJ*, 758, 24
- Frenk C. S., White S. D. M., 2012, *Annalen der Physik*, 524, 507
- Gelman A., Rubin D. B., 1992, *Statistical Science*, 7, 457
- Gelman A., Carlin J. B., Stern H. S., Rubin D. B., 2013, *Bayesian Data Analysis*, 3rd edn. Chapman and Hall/CRC, Boca Raton, FL
- Gilks W. R., 1995, *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics*, 1st edn. Chapman and Hall/CRC
- Gnedin N. Y., 2000, *ApJ*, 542, 535
- Gómez F. A., Coleman-Smith C. E., O'Shea B. W., Tumlinson J., Wolpert R. L., 2012, *ApJ*, 760, 112
- Gómez F. A., Coleman-Smith C. E., O'Shea B. W., Tumlinson J., Wolpert R. L., 2013, *arXiv:1311.2587*
- Grubbs F., 1969, *Technometrics*, 11, 1
- Guo Q., et al., 2011, *MNRAS*, 413, 101
- Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., 2003, *MNRAS*, 338, 903
- Henriques B. M. B., Thomas P. A., 2010, *MNRAS*, 403, 768
- Henriques B. M. B., Thomas P. A., Oliver S., Roseboom I., 2009, *MNRAS*, 396, 535
- Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R. E., Guo Q., Lemson G., Springel V., 2013, *MNRAS*, 431, 3373
- Hinshaw G., et al., 2013, *ApJS*, 208, 19
- Jiang F., van den Bosch F. C., 2013, *arXiv:1311.5225*
- Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, *ApJ*, 675, 1095
- Jiang L., Helly J. C., Cole S., Frenk C. S., 2014, *MNRAS*, 440, 2115
- Keisler R., et al., 2011, *ApJ*, 743, 28
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2
- Kitayama T., Suto Y., 1996, *ApJ*, 469, 480
- Kitzbichler M. G., White S. D. M., 2007, *MNRAS*, 376, 2
- Knebe A., et al., 2011, *MNRAS*, 415, 2293
- Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
- Lagos C. d. P., Lacey C. G., Baugh C. M., 2013, *MNRAS*, 436, 1787
- Lawrence E., Heitmann K., White M., Higdon D., Wagner C., Habib S., Williams B., 2010, *ApJ*, 713, 1322
- Leauthaud A., et al., 2012, *ApJ*, 744, 159
- Li C., White S. D. M., 2009, *MNRAS*, 398, 2177
- Lilly S. J., Peng Y., Carollo M., Renzini A., 2013, in *The Intriguing Life of Massive Galaxies*, Proceedings of the International Astronomical Union, IAU Symposium. eprint: *arXiv:1302.4450*, pp 141–150, doi:10.1017/S1743921313004535, <http://adsabs.harvard.edu/abs/2013IAUS...295..141L>
- Lu Y., Mo H. J., Weinberg M. D., Katz N., 2011, *MNRAS*, 416, 1949
- Lu Y., Mo H. J., Katz N., Weinberg M. D., 2012, *MNRAS*, 421, 1779
- Lu Y., Mo H. J., Lu Z., Katz N., Weinberg M. D., 2013, *arXiv:1311.0047*
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, 723, 1359
- Martin C. L., Shapley A. E., Coil A. L., Kornei K. A., Bundy K., Weiner B. J., Noeske K. G., Schiminovich D., 2012, *ApJ*, 760, 127
- Moustakas J., et al., 2013, *ApJ*, 767, 50
- Murray N., Quataert E., Thompson T. A., 2005, *ApJ*, 618, 569
- Mutch S. J., Poole G. B., Croton D. J., 2013, *MNRAS*, 428, 2001
- Neistein E., Weinmann S. M., 2010, *MNRAS*, 405, 2717
- Okamoto T., Gao L., Theuns T., 2008, *MNRAS*, 390, 920
- Oppenheimer B. D., Davé R., Kereš D., Fardal M., Katz N., Kollmeier J. A., Weinberg D. H., 2010, *MNRAS*, 406, 2325
- Padmanabhan N., et al., 2008, *ApJ*, 674, 1217
- Padmanabhan N., Xu X., Eisenstein D. J., Scalzo R., Cuesta A. J., Mehta K. T., Kazin E., 2012, *MNRAS*, 427, 2132
- Parkinson H., Cole S., Helly J., 2008, *MNRAS*, 383, 557
- Peacock J. A., Dodds S. J., 1996, *MNRAS*, 280, L19
- Percival W. J., et al., 2007, *ApJ*, 657, 645
- Planck Collaboration et al., 2013, *arXiv:1303.5076*
- Riess A. G., et al., 2011, *ApJ*, 730, 119
- Ruiz A. N., et al., 2013, *arXiv:1310.7034*
- Smith R. E., 2012, *MNRAS*, 426, 531
- Springel V., et al., 2005, *Nature*, 435, 629
- Srisawat C., et al., 2013, *MNRAS*
- Stefansky W., 1972, *Technometrics*, 14, 469

- Sullivan M., et al., 2011, *ApJ*, 737, 102
 Taylor E. N., et al., 2011, *MNRAS*, 418, 1587
 Terr Braak C. J. F., 2006, *Stat Comput*, 16, 239
 Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
 Weinberg M. D., 2013, *MNRAS*, 434, 1736

APPENDIX A: CONSTRAINTS ON HOD MODEL PARAMETERS

In computing the covariance matrix of the Li & White (2009) stellar mass function we make use of a parametric form for the HOD given by Behroozi et al. (2010). The reader is referred to that paper for a full specification of the parametric form (see also Appendix B). We constrain 11 parameters in the HOD using the BIE, using priors as given in Table A1.

We use the differential evolution algorithm of Terr Braak (2006) as our MCMC sampler, with $\gamma(\equiv 0.1\gamma_0) = 0.0507$ where $\gamma_0 = 2.38/\sqrt{N_{\text{dim}}}$ is the optimal fraction for a multidimensional Gaussian posterior (Terr Braak 2006) and N_{dim} is the dimension of the parameter space. This results in a reasonable acceptance rate in the chains. Every 10 steps, we set $\gamma = 1$ to allow chains to swap modes.

Technically, to ensure that the chains are positively recurrent in the differential evolution algorithm it is necessary to add a random element to each proposed step in the MCMC algorithm (Terr Braak 2006). This random element should have small variance, but long tails to large values. For this purpose we draw the random components of the proposed step from Cauchy distributions with widths, ϵ , as given in Table A1.

We employ a Gelman-Rubin \hat{R} statistic (Gelman & Rubin 1992) to determine when the chains have converged. When $\hat{R} \leq 1.2$ we declare convergence and allow the chains to run for a further 10^6 steps. Only these final 10^6 steps are used to assess the posterior probability distribution of the HOD parameters. The final posterior probability distribution is shown in Figure A1.

Finally, we take the maximum likelihood solution as our choice of parameters for the HOD and use these in computing the covariance matrix of the Li & White (2009) stellar mass function. This, of course, throws away all information about the posterior probability distribution of the HOD parameters that the BIE has determined. One could imagine instead sampling parameter sets from the posterior probability distribution and building a sample of covariance matrices representative of the uncertainties in the HOD parameters. When constructing the likelihood function for our galaxy formation model one could then include this additional covariance in the data covariance matrix. Given the rather strong constraints on the HOD parameters we do not feel that the substantial extra complexity and computation needed to achieve this is warranted, although we intend to revisit this point in future.

APPENDIX B: OPTIMAL SAMPLING OF $Z = 0$ HALO MASSES

We wish to answer the following question: given some finite amount of computing time, what is the optimal distribution of halo masses to run when comparing to a given dataset? For example, is it better to run a volume limited sample (as one would get from an N-body simulation) or is it better to use equal numbers of halos per logarithmic interval of halo mass? The following section describes how to solve this optimization problem in the specific case of fitting to the stellar mass function.

We define:

$n(M)d \ln M$ is the dark matter halo mass function, i.e. the number of halos in the range M to $M + M d \ln M$ per unit volume;

$\gamma(M)d \ln M$ is the number of trees that we will simulate in the range M to $M + M d \ln M$;

$\alpha(M_*)$ is the error on the observed stellar mass function at mass M_* ;

$P(N|M_*, M; \delta \ln M_*)$ is the conditional stellar mass distribution function of galaxies of stellar mass M_* in a bin of width $\delta \ln M_*$ per halo of mass M ;

$t(M)$ is the CPU time it takes to simulate a tree of mass M .

To clarify, $P(N|M_*, M; \delta \ln M_*)$ is the probability³⁷ to find N galaxies of mass between M_* and $M_* + M_* \delta \ln M_*$ in a halo of mass M . The usual conditional stellar mass function is simply the first moment of this distribution:

$$\phi(M_*; M) \delta \ln M_* = \sum_{N=0}^{\infty} N P(N|M_*, M; \delta \ln M_*). \quad (\text{B1})$$

The model estimate of the stellar mass function $\Phi(M_*)$ (defined per unit $\ln M_*$) is

$$\Phi(M_*) = \int_0^{\infty} \phi(M_*; M) \frac{n(M)}{\gamma(M)} \gamma(M) d \ln M, \quad (\text{B2})$$

where the $n(M)/\gamma(M)$ term is the weight assigned to each tree realization—and therefore the weight assigned to each model galaxy when summing over a model realization to construct the stellar mass function.

When computing a model likelihood, we must employ some statistic which defines how likely the model is given the data. Typically, for stellar mass functions we have an estimate of the variance in the data, $\alpha^2(M_*)$, as a function of stellar mass (the full covariance matrix could easily be incorporated into this method if necessary). In that case, we can define a likelihood

$$\ln \mathcal{L} = -\frac{1}{2} \sum_i \frac{[\phi_{\text{obs},i} - \phi_i]^2}{\alpha_i^2 + \sigma_i^2} \quad (\text{B3})$$

where the sum is taken over all data points, i , and σ_i^2 is the variance in the model estimate and is given by

$$\sigma^2(M_*) = \langle [\phi(M_*) - \bar{\phi}(M_*)]^2 \rangle, \quad (\text{B4})$$

where $\phi(M_*)$ is the realization from a single model and $\bar{\phi}(M_*)$ is the model expectation from an infinite number

³⁷ To put it another way, $P(N|M_*, M; \delta \ln M_*)$ is closely related to the commonly used Halo Occupation Distribution.

Table A1. Adopted priors for parameters of our HOD model. For α_{sat} , the slope of the satellite HOD at high masses, we adopt a prior consistent with the results of Kravtsov et al. (2004). For all other parameters we adopted uniform priors spanning a wide range. Also shown are the widths, ϵ , of the Cauchy distributions used to add a random component to each proposed step in our chains.

Parameter	Units	Prior	ϵ
α_{sat}	—	$N(0.98, 7.55 \times 10^{-4}, -\infty, +\infty)$	0.0002929
M_1	M_\odot	$U_{\ln}(12.1, 12.6)$	0.0005773
$M_{*,0}$	M_\odot	$U_{\ln}(10.5, 10.8)$	0.0003486
β	—	$U(0.35, 0.5)$	0.0002719
δ	—	$U(0.4, 0.65)$	0.0003668
γ	—	$U(0.7, 1.9)$	0.001382
$\sigma_{\log M_*}$	—	$U(0.1, 0.42)$	0.0001321
B_{cut}	—	$U(1.0, 128.0)$	0.008308
B_{sat}	—	$U(1.0, 20.0)$	0.01833
β_{cut}	—	$U(-2.0, 0.0)$	0.003111
β_{sat}	—	$U(1.0, 2.0)$	0.02055

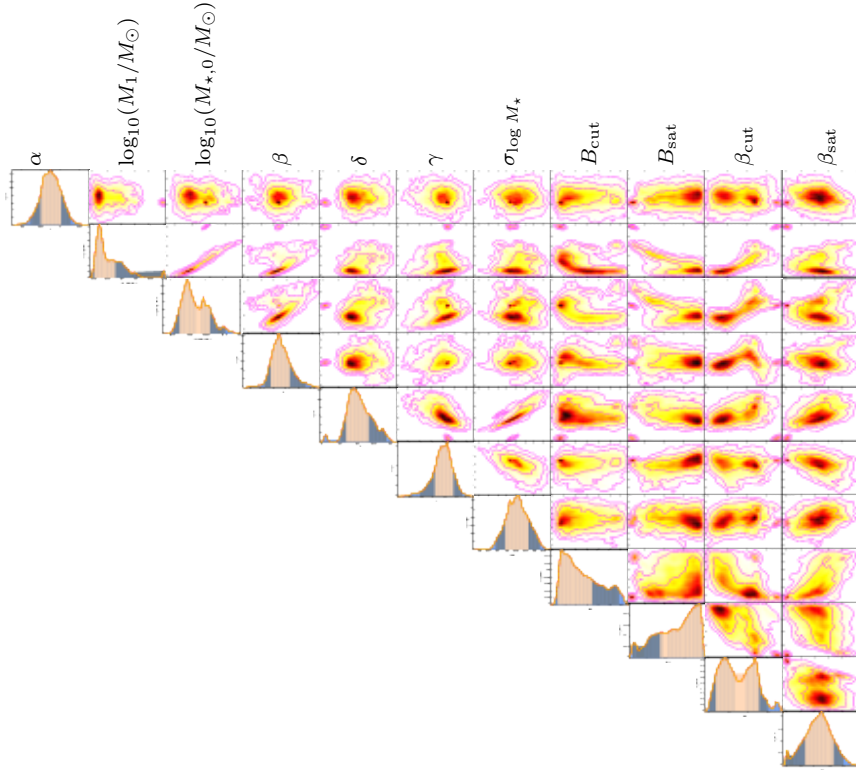


Figure A1. The posterior probability distribution of parameters in our parametric HOD constrained to fit the observed stellar mass function of Li & White (2009). The results shown are for the third cycle of our iterative procedure in which we constrain the HOD parameters using a covariance matrix derived from the previous set of HOD parameters. Panels on the diagonal show the marginalized posterior probability distribution for each individual parameter, while off-diagonal panels show the posterior probability distribution for pairs of parameters, marginalized over all other parameters.

of merger tree realizations and the average is taken over all possible model realizations. Since the contributions from each merger tree are independent,

$$\sigma^2(M_*) = \sum_i \zeta_i^2(M_*; M) \quad (\text{B5})$$

where $\zeta_i^2(M_*; M)$ is the variance in the contribution to the stellar mass function from tree i . This in turn is given by

$$\zeta_i^2(M_*; M) = \psi^2(M_*; M) \left[\frac{n(M)}{\gamma(M)} \right]^2, \quad (\text{B6})$$

where $\psi^2(M_*; M)$ is the variance in the conditional stellar mass function. In the continuum limit this becomes

$$\sigma^2(M_*) = \int_0^\infty \psi^2(M_*; M) \left[\frac{n(M)}{\gamma(M)} \right]^2 \gamma(M) d \ln M. \quad (\text{B7})$$

Model variance artificially increases the likelihood of a given model. We would therefore like to minimize the increase in the likelihood due to the model variance:

$$\Delta 2 \ln \mathcal{L} = \sum_i \frac{[\phi_{\text{obs},i} - \phi_i]^2}{\alpha_i^2} - \frac{[\phi_{\text{obs},i} - \phi_i]^2}{\alpha_i^2 + \sigma_i^2} \quad (\text{B8})$$

Of course, we don't know the model prediction, ϕ_i , in advance³⁸. However, if we assume that a model exists which is a good fit to the data then we would expect that $[\phi_{\text{obs},i} - \phi_i]^2 \approx \alpha_i^2$ on average. In that case, the increase in likelihood due to the model is minimized by minimizing the function³⁹

$$F[\gamma(M)] = \sum_i \frac{\alpha_i^2}{\alpha_i^2 + \sigma_i^2}. \quad (\text{B9})$$

If the bins all have the same $\delta \ln M_*$ we can turn the sum into an integral

$$F[\gamma(M)] = \int_0^\infty \frac{\alpha(M_*)^2}{\alpha(M_*)^2 + \sigma(M_*)^2} d \ln M_*. \quad (\text{B10})$$

Obviously, the answer is to make $\gamma(M) = \infty$, in which case $F[\gamma(M)] = 0$. However, we have finite computing resources. The total time to run our calculation is

$$\tau = \int_0^\infty t(M) \gamma(M) d \ln M. \quad (\text{B11})$$

We therefore want to minimize $F[\gamma(M)]$ while keeping τ equal to some finite value. We can do this using a Lagrange multiplier and minimizing the function

$$F[\gamma(M)] = \int_0^\infty \frac{\alpha(M_*)^2}{\alpha(M_*)^2 + \sigma(M_*)^2} d \ln M_* + \int_0^\infty \lambda \gamma(M) t(M) d \ln M. \quad (\text{B12})$$

Finding the functional derivative and setting it equal to zero gives:

$$\gamma(M) = \sqrt{\frac{\xi(M)}{\lambda t(M)}}, \quad (\text{B13})$$

in the limit where⁴⁰ $\sigma(M_*) \ll \alpha(M_*)$, and where

$$\xi(M) = n^2(M) \int_{-\infty}^\infty \frac{\psi^2(M_*; M)}{\alpha^2(M_*)} d \ln M_*. \quad (\text{B14})$$

The values of λ and $\delta \ln M_*$, and the normalization of $t(M)$ are unimportant here since we merely want to find the optimal shape of the $\gamma(M)$ function—we can then scale it up or down to use the available time.

Figure B3 shows the function $\gamma(M)$ obtained by adopting a model conditional stellar mass function which is a sum of central and satellite terms. Specifically, we use the model of Leauthaud et al. (2012) which is constrained to match observations from the COSMOS survey. In their model⁴¹:

$$\begin{aligned} \langle N_c(M_*|M) \rangle &\equiv \int_{M_*}^\infty \phi_c(M'_*) d \ln M'_* \\ &= \frac{1}{2} \left[1 - \text{erf} \left(\frac{\log_{10} M_* - \log_{10} f_{\text{SHMR}}(M)}{\sqrt{2} \sigma_{\log M_*}} \right) \right]. \end{aligned} \quad (\text{B15})$$

³⁸ Below, we will adopt a simple empirical model for $\phi(M_*)$. However, it should not be used here since we will in actuality be computing the likelihood from the model itself.

³⁹ This can be seen intuitively: we are simply requiring that the variance in the model prediction is small compared the the variance in the data.

⁴⁰ This is the limit in which we would like our results to be.

⁴¹ This integral form of the conditional stellar mass function is convenient here since it allows for easy calculation of the number of galaxies expected in the finite-width bins of the observed stellar mass function.

Table B1. Parameters of the conditional stellar mass function fit.

Parameter	Value
α_{sat}	1.0
$\log_{10} M_1$	12.120
$\log_{10} M_{*,0}$	10.516
β	0.430
δ	0.5666
γ	1.53
$\sigma_{\log M_*}$	0.206
B_{cut}	0.744
B_{sat}	8.00
β_{cut}	-0.13
β_{sat}	0.859

Here, the function $f_{\text{SHMR}}(M)$ is the solution of

$$\log_{10} M = \log_{10} M_1 + \beta \log_{10} \left(\frac{M_*}{M_{*,0}} \right) + \frac{(M_*/M_{*,0})^\delta}{1 + (M_*/M_{*,0})^{-\gamma}} - 1/2. \quad (\text{B16})$$

For satellites,

$$\begin{aligned} \langle N_s(M_*|M) \rangle &\equiv \int_{M_*}^\infty \phi_s(M'_*) d \ln M'_* \\ &= \langle N_c(M_*|M) \rangle \left(\frac{f_{\text{SHMR}}^{-1}(M_*)}{M_{\text{sat}}} \right)^{\alpha_{\text{sat}}} \\ &\quad \times \exp \left(- \frac{M_{\text{cut}}}{f_{\text{SHMR}}^{-1}(M_*)} \right), \end{aligned} \quad (\text{B17})$$

where

$$\frac{M_{\text{sat}}}{10^{12} M_\odot} = B_{\text{sat}} \left(\frac{f_{\text{SHMR}}^{-1}(M_*)}{10^{12} M_\odot} \right)^{\beta_{\text{sat}}}, \quad (\text{B18})$$

and

$$\frac{M_{\text{cut}}}{10^{12} M_\odot} = B_{\text{cut}} \left(\frac{f_{\text{SHMR}}^{-1}(M_*)}{10^{12} M_\odot} \right)^{\beta_{\text{cut}}}. \quad (\text{B19})$$

We use the best fit parameters from the SIGMOD1 method of Leauthaud et al. (2012) for their z_1 sample, but apply a shift of -0.2 dex in masses to bring the fit into line with the $z = 0.07$ mass function of Li & White (2009). The resulting parameter values are shown in Table B1.

We assume that $P_s(N|M_*, M; \delta \ln M_*)$ is a Poisson distribution while $P_c(N|M_*, M; \delta \ln M_*)$ has a Bernoulli distribution, with each distribution's free parameter fixed by the constraint of eqn. (B1), and the assumed forms for ϕ_c and ϕ_s .

The errors in the Li & White (2009) observed stellar mass function are well fit by (see Fig. B1):

$$\begin{aligned} \alpha(M_*) &= 10^{-3} \left(\frac{M_*}{4.5 \times 10^{10} M_\odot} \right)^{-0.3} \exp \left(- \frac{M_*}{4.5 \times 10^{10} M_\odot} \right) \\ &\quad + 10^{-7}, \end{aligned} \quad (\text{B20})$$

and the tree processing time in our model can be described by:

$$\log_{10} t(M) = \sum_{i=0}^2 C_i [\log_{10} M]^i \quad (\text{B21})$$

with $C_0 = -16.25$, $C_1 = 1.767$ and $C_2 = 0.02883$ (see the left-hand panel of Fig. B2).

The resulting optimal sampling density curve is shown in Fig. B3 and is compared to weighting by the halo mass function (i.e. the result of sampling halos at random from a representative volume). Optimal sampling gives less weight to low mass halos (since a sufficient accuracy can be obtained without the need to run many tens of thousands of such halos) and to high mass halos which are computationally expensive. Nevertheless, the optimal sampling curve is quite close to the halo mass function, so we would expect that optimal sampling and sampling from the halo mass function should be similarly efficient. The right-hand panel of Fig. B2 shows the convergence measure $F[\gamma(M)]$ as a function of total tree processing time for three different methods of sampling the halo mass function. The optimal method achieves the smallest convergence measure in a given total processing time. It is clear that “haloMassFunction” (i.e. sampling halos in proportion to the halo mass function; green points) and “stellarMassFunction” (i.e. using the optimal weighting; red points) samplings give almost equally efficient results⁴². Sampling using “powerLaw” ($\gamma(M) \propto [\log_{10} M]^{-1/2}$; blue points) is much less efficient.

Clearly in this case our optimal sampling method did not improve over simply sampling from the halo mass function, although it does almost equally as well. This is likely because calculation of the optimal sampling requires a model for the conditional mass function of halos (eqn. B1). If this model does not match the conditional mass function in our actual SAM then the computed sampling will not be optimal. Once a viable SAM is obtained it can be used to directly compute the conditional mass function, which should allow this method to more accurately compute the optimal sampling.

Given these results, for the present work, we choose to use sampling from the halo mass function in our MCMC analysis⁴³. Additionally, we introduce upper and lower limits on the halo abundance, *for the purposes of sampling halo mass only*, to ensure that a minimum number of high mass halos, and not too many low mass halos are simulated.

APPENDIX C: ROBUSTNESS AND CONVERGENCE OF MODEL RESULTS

For the purposes of these robustness and convergence tests (and for the computation of model discrepancy terms; see §3.5) we adopt an “*a priori*” set of model parameters which were found to give a reasonable match (by eye, not in any strict statistical sense) to the observed stellar mass function.

⁴² The numbers next to each point indicate the number of merger trees run on average per decade of halo mass. It can be seen that the “stellarMassFunction” sampling achieves the same convergence measure as “haloMassFunction” sampling using fewer merger trees. However, because “stellarMassFunction” sampling tends to sample more high mass trees, the total time required is about the same.

⁴³ Note that the optimal sampling function here was computed assuming a fixed mass resolution in merger trees. In our MCMC analysis we instead use a variable mass resolution. We have computed an optimal sampling for that case also, and find that it performs equally well, but not better than sampling from the halo mass function.

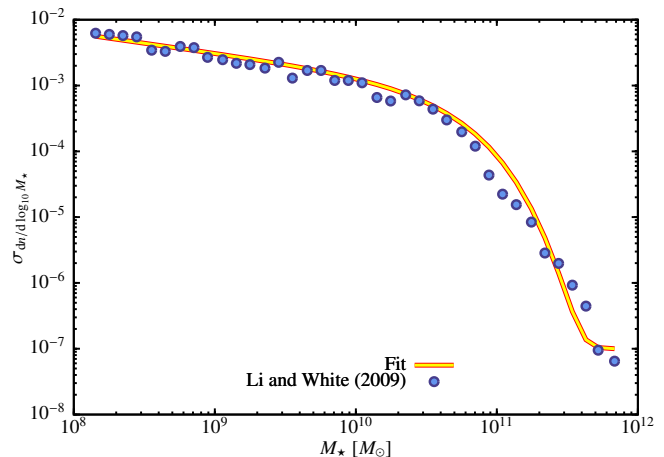


Figure B1. Errors on the Li & White (2009) stellar mass function (points) and the fitting function (line) given by eqn. (B20).

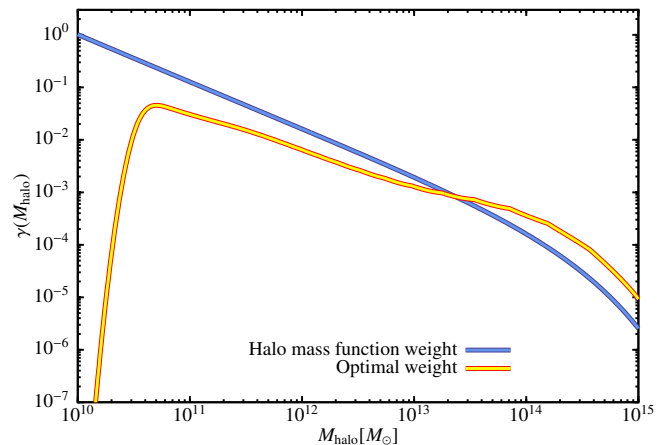


Figure B3. Optimal weighting ($\gamma(M)$; yellow line) compared with weighting by the dark matter halo mass function (i.e. sampling halos at random from a representative volume; blue line). Sampling densities have been normalized to unit compute time.

Obviously, prior to constraining our model to match the observational data we do not know the maximum likelihood set of parameters, and so must make a “best guess” for the purposes of evaluating these models—our “*a priori*” model serves this purpose. Parameters of the “*a priori*” model are given in Table C1.

Before beginning our MCMC analysis we test that our model is converged with respect to all numerical parameters. To test for convergence, we run an *a priori* model multiple times, each time adjusting the value of a single numerical parameter, and compute the model stellar mass function (i.e. precisely the same statistic that we will use in our MCMC study). Thereby, we generate a set of model stellar mass functions as a function of the values of model numerical parameters. For each sequence of models, in which the value of a single numerical parameter is varied, we define the *most optimal* model as that with the lowest or highest value of the model parameter, depending on whether the most accurate results are expected for small or large values of the parameter.

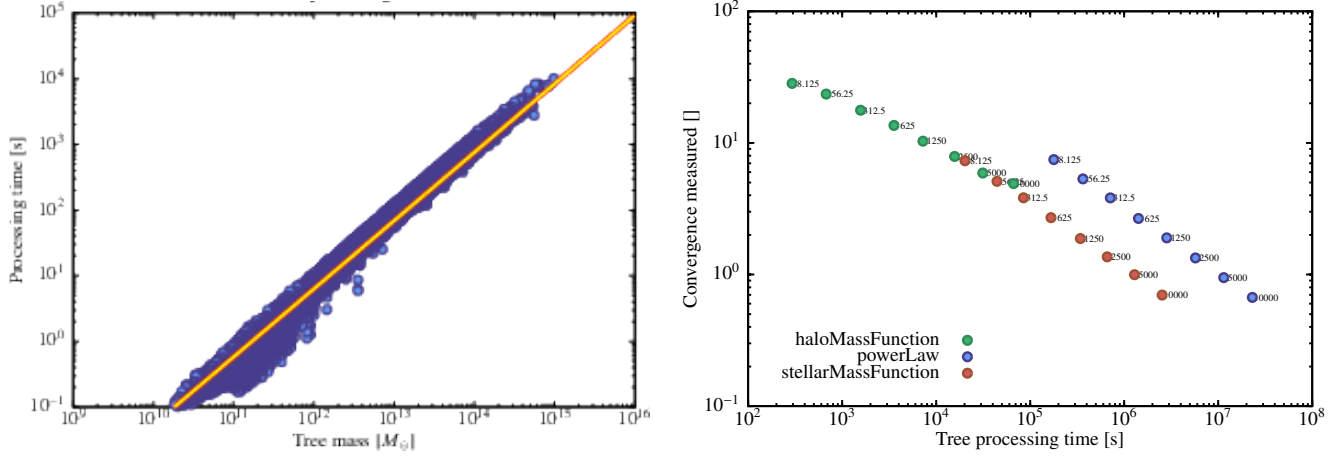


Figure B2. *Left panel:* The CPU time taken to a merger tree of given $z = 0$ mass. Blue points show measured values for a large number of model merger trees. The yellow line indicates the best-fit 2nd-order polynomial to this relation as described in eqn. (B21). *Right panel:* The convergence measure, defined in eqn. (B9), as a function of total time taken to process the set of merger trees. Results are shown for three different choices for sampling masses of present day dark matter halos as described in the text.

Parameter	Value
α_{cool}	0.35
β_{cool}	0.50
$\Delta \log_{10} \mathcal{M}_{\text{cool}}$	0.30
τ_{cool}	2.26 Gyr
$\mathcal{M}_{\text{cool}}$	$3.0 \times 10^{11} M_{\odot}$
α_{\star}	0.21
β_{\star}	2.40
$\tau_{\star, \text{min}}$	3.50 Myr
τ_{\star}	0.68 Gyr
α_{wind}	-6.68
β_{wind}	4.78
f_{wind}	0.0138
$\tau_{\text{wind}, \text{min}}$	5.09 Myr
f_{df}	8.35
V_{reion}	46.69 km/s
$C_{1...6}$	0.00
μ	-0.27
κ	0.0127
H_1	-0.0198
H_2	0.00986

Table C1. Parameters of the “*a priori*” model used for tests of model robustness and accuracy, and also for calculations of model discrepancy terms.

We then define a convergence measure

$$\chi^2 = \frac{\Delta \cdot [\mathcal{C}_{(\text{model})} + \mathcal{C}_{(\text{model}, \text{opt})}]^{-1} \cdot \Delta^T}{N_{\text{eff}}} \quad (\text{C1})$$

where Δ is the difference between a model evaluation of the stellar mass function and the that obtained using the same model but with the most optimal value of the parameter being tested, $\mathcal{C}_{(\text{model})}$ is the covariance matrix of the model, $\mathcal{C}_{(\text{model}, \text{opt})}$ is the covariance matrix of the optimal model, and N_{eff} is an effective number of degrees of freedom. This measure should approach unity as convergence is achieved. The error on this convergence measure is

$$\frac{\sigma_{\chi^2}^2}{\chi^2} = \frac{2}{N_{\text{eff}}} + \left(\frac{\sigma_{N_{\text{eff}}}}{N_{\text{eff}}} \right)^2. \quad (\text{C2})$$

The effective number of degrees of freedom is found by running many realizations of the “*a priori*” model with different random seeds but otherwise unchanged parameters and taking the average of $\Delta \cdot [\mathcal{C}_{(\text{model})} + \mathcal{C}_{(\text{model}, \text{opt})}]^{-1} \cdot \Delta^T$ over all realizations. This is necessary to account for the effects of significant multicollinearity in the model covariance matrix which result in this quantity being distributed as a χ^2 -distribution with fewer degrees of freedom than would be expected.

Figures C1 and C2 show convergence diagrams for all numerical parameters in our model. The solid horizontal line shows $\chi^2 = 1$ (the expected value for converged models), the red point (shown at arbitrary χ^2) indicates the value of the parameter used in our models for the MCMC analysis, the green arrow indicates the direction of increasing optimality of the parameter (i.e. moving in the direction of the arrow should lead to better converged models), and the blue points indicate the convergence measures obtained by running models while varying the parameter shown on the x -axis.

The top row of Fig. C1 shows convergence in the parameters `mergerTreeBuildHaloMassMinimum` and `mergerTreeBuildHaloMassMaximum` which control the minimum and maximum masses halos at $z = 0$ that are used as the roots of merger trees. The range spanned must be sufficient to capture all galaxies which contribute significantly to the stellar mass function. Clearly for both parameters the model is converged for our choice of values. In the case of `mergerTreeBuildHaloMassMinimum` the model remains converged up to $2 \times 10^{11} M_{\odot}$ (the highest value we consider). Nevertheless, we retain a value of $2 \times 10^8 M_{\odot}$ for this parameter since, as the MCMC algorithm explores parameter space it can sometimes produce models which form galaxies much more efficiently in lower mass halos, thereby requiring a lower value of `mergerTreeBuildHaloMassMinimum` to achieve convergence. Given the available mass of baryons in $2 \times 10^8 M_{\odot}$ halos, this value of `mergerTreeBuildHaloMassMinimum` is sufficient for convergence in any region of parameter space that the MCMC algorithm might explore. For

`mergerTreeBuildHaloMassMaximum`, convergence is achieved above $6 \times 10^{14} M_{\odot}$ —higher mass trees are sufficiently rare that they do not contribute significantly to the stellar mass function.

The second and third rows of Fig. C1 shows convergence in parameters which control the size of the timestep taken when constructing merger trees using the algorithm of Parkinson et al. (2008). The parameters `mergerTreeBuildCole2000AccretionLimit` and `mergerTreeBuildCole2000MergeProbability` correspond to the maximum allowed values of the quantities that Parkinson et al. (2008) label F and P (their eqns. 4 and 5). Model results are clearly well-converged with respect to these parameters. The parameter `modifiedPressSchechterFirstOrderAccuracy` corresponds to ϵ_1 defined by Parkinson et al. (2008; their Appendix A) and also limits timesteps in tree building⁴⁴. Convergence in this parameter is just barely reached at our chosen value. Finally, the bottom row of Fig. C1 shows parameters controlling the mass resolution of merger trees. Parameters `mergerTreeBuildMassResolutionScaledMinimum` and `mergerTreeBuildMassResolutionScaledFraction` correspond to $M_{\text{res,min}}$ and f_0 of eqn. (1) respectively. `mergerTreeBuildMassResolutionScaledMinimum` is well-converged at our chosen value, but `mergerTreeBuildMassResolutionScaledFraction` is clearly far from being converged at our chosen value of 0.1. However, we are forced to adopt this large value to permit models to be computed sufficiently rapidly. Knowing that this will cause of models to give discrepant answers, we include a model discrepancy term to account for this approximation.

Figure C2 shows convergence diagrams for six parameters which control time-stepping in our model. The top row shows `odeToleranceAbsolute` and `odeToleranceRelative`. In our model the ODE solver uses adaptive timesteps to keep the error in variable i below

$$\delta_i = \epsilon_{\text{abs}} s_i + \epsilon_{\text{rel}} |y_i|, \quad (\text{C3})$$

where y_i is the value of the i^{th} variable, s_i is a fixed value⁴⁵ for the i^{th} variable, and ϵ_{abs} and ϵ_{rel} correspond to `odeToleranceAbsolute` and `odeToleranceRelative` respectively. The model is clearly well-converged with respect to both of these parameters.

Nodes of merger trees in our model are evolved individually—that is, each node will evolve for some time before the code moves on to another node. This means that nodes are always slightly out-of-sync (they are, of course, brought in to sync at each requested output time). If galaxies in nodes are non-interacting this has no effect on the overall evolution of the tree and its constituent galaxies. However, galaxies do interact—specifically satellite galaxies interact with their host halo (and the central galaxy of that host) by transferring their hot gas component to the host (“strangulation”) and by merging with the central galaxy. It is important therefore to prevent satellites and hosts from

getting too far out of sync. The four parameters shown in the lower two rows of Fig. C2 limit how far satellites can get out of sync with their hosts before their evolution is halted until the host node “catches up”. The reader is referred to Benson (2012) for full details of these parameters. The convergence diagrams for these parameters controlling time-stepping of satellites and hosts all clearly show that the model is converged with respect to these parameters.

Convergence potentially depends on the parameters of the model being varied in the MCMC analysis. Therefore, once the MCMC analysis was completed we repeated this convergence study using the maximum likelihood set of parameters. The convergence characteristics of most parameters remain unchanged. There are two cases where the model is less well converged when using the maximum likelihood model compared to the *a priori* model, both of which affect only the high-mass end of the galaxy stellar mass function. The first case is for the parameter `modifiedPressSchechterFirstOrderAccuracy` which controls merger tree construction. With the *a priori* model, this parameter was just barely converged at the value adopted in this work. With the maximum likelihood model this is no longer true, and `modifiedPressSchechterFirstOrderAccuracy` would need to be decreased by a factor of around 3 to restore convergence. The second case is the timestepping parameters `timestepHostRelative` and `timestepHostAbsolute`. These were set to 0.1 and 1 Gyr respectively in this work. For the maximum likelihood model to be converged would require either `timestepHostRelative` to be reduced to 0.003, or for `timestepHostAbsolute` to be reduced to 0.03 Gyr. This change in convergence characteristics reflects differences in the timescales for star formation and outflows in galaxies in the maximum likelihood and *a priori* models.

We note that, while we can detect this lack of convergence in our model by running simulations using a very large number of merger trees (to reduce statistical noise), the resulting changes in the galaxy stellar mass function are small compared to the statistical noise in the models that we actually run in our MCMC analysis (which utilize many fewer merger trees to keep calculation times sufficiently short).

Nevertheless, this suggests that, in future work, an iterative approach be adopted (see also §4.3.4 where a similar approach is recommended for model discrepancy calculations), in which convergence is tested for the current maximum likelihood model, new constraints are introduced, the model parameter space is explored quickly using MCMC, and convergence is re-tested using the new maximum likelihood model. This process can be iterated until convergence characteristics are no longer changing.

APPENDIX D: MODEL MASS FUNCTION COVARIANCE MATRIX MODEL

Our SAM uses a Monte Carlo approach to generate results, simulating a random sample of dark matter merger trees and then summing over their constituent galaxies. As such, the results from the model always have some random fluctuation around the “true” answer for a given set of input parameters. This additional covariance should be accounted for when evaluating model likelihoods. Ideally, the covariance would

⁴⁴ Specifically, it ensures that eqn. (2) of Parkinson et al. (2008) is first-order accurate.

⁴⁵ More specifically, s_i is fixed over a given timestep. It may change between timesteps due to possible dependencies on other variables.

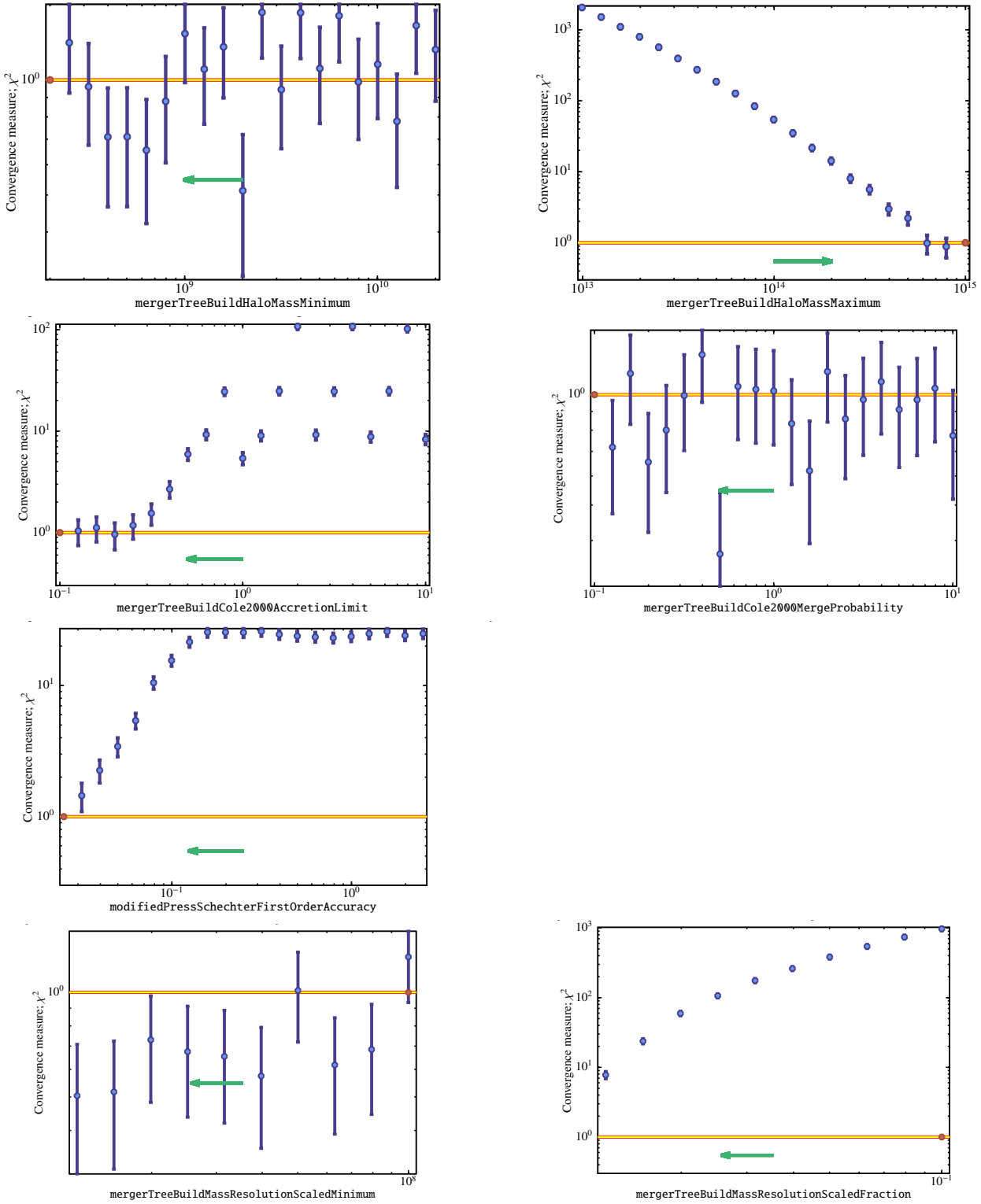


Figure C1. Convergence diagrams for model parameters that control the construction of dark matter merger trees. The solid horizontal lines show $\chi^2 = 1$ (the expected value for converged models), the red point (shown at arbitrary χ^2) indicates the value of the parameter used in our models for the MCMC analysis, the green arrow indicates the direction of increasing optimality of the parameter (i.e. moving in the direction of the arrow should lead to better converged models), and the blue points indicate the convergence measures obtained by running models while varying the parameter shown on the x -axis.

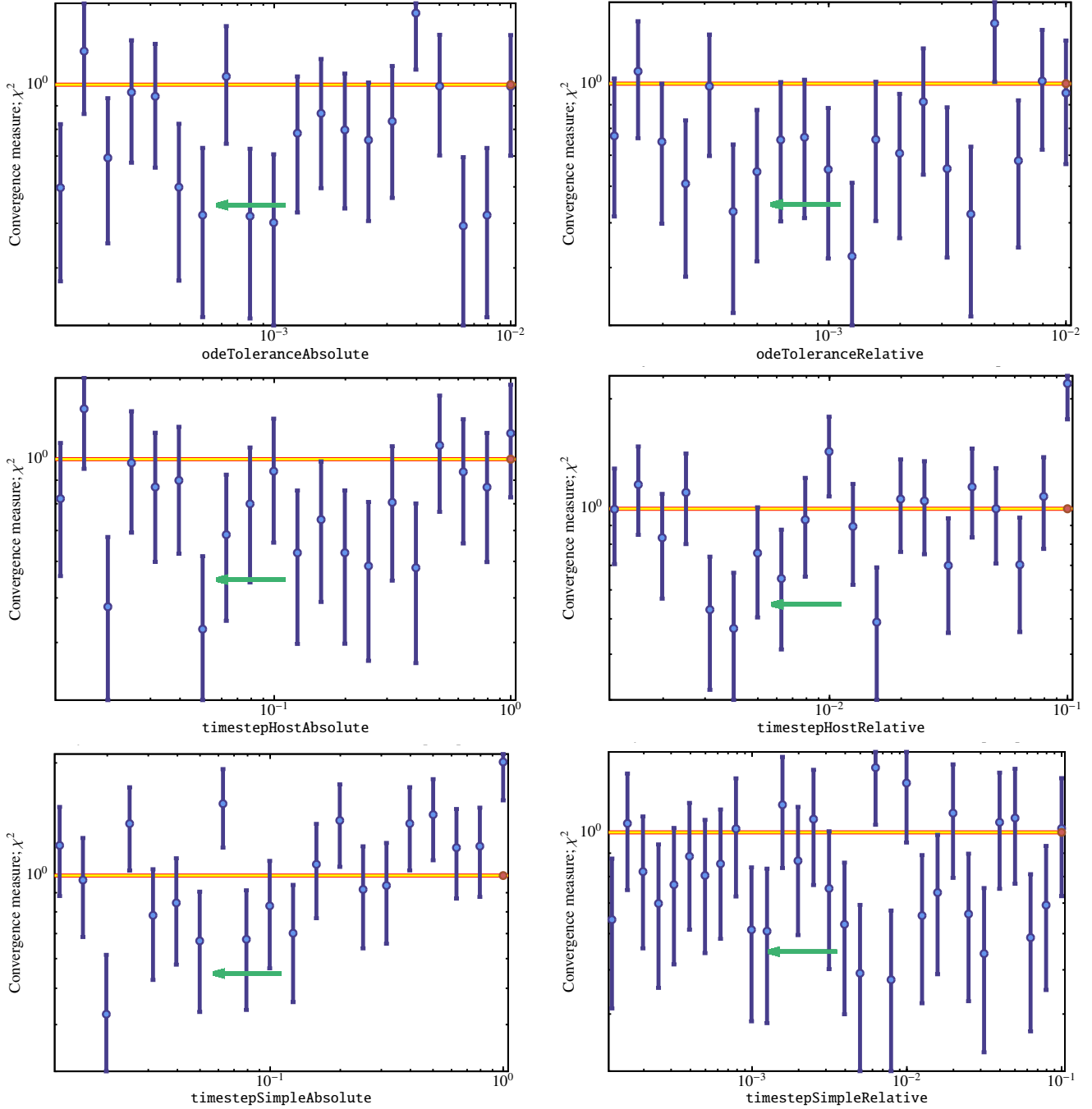


Figure C2. Convergence diagrams for model parameters that control the time-stepping when evolving systems of ODEs through merger trees. The solid horizontal lines show $\chi^2 = 1$ (the expected value for converged models), the red point (shown at arbitrary χ^2) indicates the value of the parameter used in our models for the MCMC analysis, the green arrow indicates the direction of increasing optimality of the parameter (i.e. moving in the direction of the arrow should lead to better converged models), and the blue points indicate the convergence measures obtained by running models while varying the parameter shown on the x -axis.

be made so small (by generating a sufficiently large number of Monte Carlo realizations) that it is entirely dominated by the covariance arising from the observational data, and so could be ignored. In practice, the MCMC technique is so time consuming that we are always driven to generate “just enough” Monte Carlo merger tree realizations to keep the covariance non-dominant, but non-negligible.

It is therefore important to have a good model for the

covariance of the model results. GALACTICUS works by sampling a set of tree “root masses” from the $z = 0$ dark matter halo mass function. From each root, a tree is grown, within which the physics of galaxy formation is then solved. Root masses are sampled uniformly from the halo mass function. That is, the cumulative halo mass function, $N(M)$, is constructed between the maximum and minimum halo masses to be simulated. The number of root masses, N_r , to be used

in a model evaluation is then determined. Root masses are then chosen such that

$$N(M_i) = N(M_{\min}) \frac{i-1}{N_r-1} \quad (\text{D1})$$

for $i = 1 \dots N_r$ (noting that $N(M_{\max}) = 0$ by construction).

Consider first those galaxies which form in the main branch of each tree (i.e. those galaxies which are destined to become the central galaxy of the $z = 0$ halo). Suppose that we simulate N_k halos of root mass M_k at $z = 0$. In such halos the main branch galaxies will, at any time, have stellar masses drawn from some distribution $p_k(M_*|t)$. The number of such galaxies contributing to bin i of the mass function is therefore binomially distributed with success probability $p_{ik} = \int_{M_{i,\min}}^{M_{i,\max}} p_k(M_*|t) dM_*$ and a sample size of N_k . The contribution to the covariance matrix from these main branch galaxies is therefore:

$$C_{ij} = \begin{cases} p_{ik}(1-p_{ik})N_k w_k^2 & \text{if } i = j, \\ -p_{ik}p_{jk}N_k w_k^2 & \text{otherwise,} \end{cases} \quad (\text{D2})$$

where w_k is the weight to be assigned to each tree. To compute this covariance requires knowledge of the probabilities, p_{ik} . We estimate these directly from the model. To do this, we bin trees into narrow bins of root mass and assume that p_{ik} does not vary significantly across the mass range of each bin. Using all realizations of trees that fall within a given bin, k , we can directly estimate p_{ik} .

In addition to the main branch galaxies, each tree will contain a number of other galaxies (these will be “satellite” galaxies at $z = 0$, but at higher redshifts may still be central galaxies in their own halos). Previous studies (Kravtsov et al. 2004) have established that the number of satellites in halos is well described by a Poisson process.

To test these assumptions we run 1000 realizations of our *a priori* model, each time using a different random seed such that each model is a statistically independent realization. For each model we construct the mass function. From this ensemble of model mass function estimates we directly measure the covariances of main branch galaxies, other galaxies, and the covariance between these two samples.

Figure D1 shows the measured covariance, and the ratio of model to measured covariance in each case. Our model (in which we assume no covariance between main branch and other galaxies) clearly performs quite well. To quantify this more precisely we construct a test statistic, \mathcal{T} , defined as:

$$\mathcal{T} = \Delta \cdot C^{-1} \cdot \Delta^T, \quad (\text{D3})$$

where Δ is the difference between the mass function of a model realization and the mean mass function of all model realizations. If our model for the model covariance matrix is accurate, the distribution of \mathcal{T} should be consistent with that obtained using the covariance matrix measured directly from the model realizations. Figure D2 shows that this is indeed the case. For comparison, we show the results if we assume that main branch galaxies follow a Poisson process. Clearly this overestimates the covariance in the model and results in an incorrect \mathcal{T} distribution.

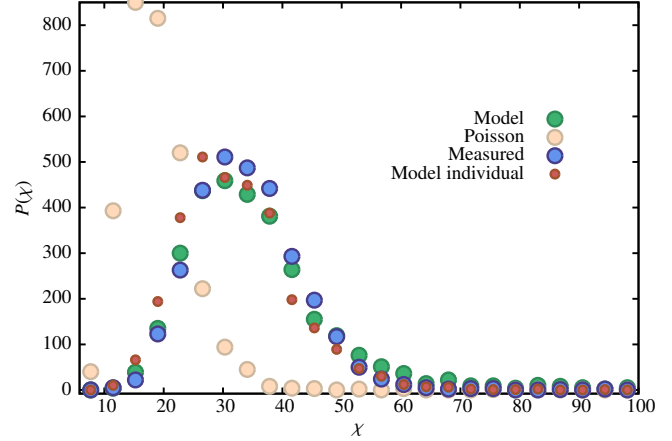


Figure D2. The test statistic, \mathcal{T} (defined by eqn. D3), for models of the model covariance matrix. Points show the distribution of test statistic obtained from 1000 model realizations. Large blue circles show \mathcal{T} obtained using the covariance matrix measured directly from the 1000 model realizations (i.e. the true covariance matrix). Large green points show the test statistic obtained using the binomial model described in the text when the resulting covariance matrix is averaged over all 1000 realizations. Small red points are the same, but the covariance matrix used for each realization is estimated from that realization alone. Finally, the large pale pink points indicate \mathcal{T} obtained when the covariance matrix is estimated assuming a Poisson process.

APPENDIX E: MERGER TREE CONSTRUCTION

Throughout this work we utilize merger trees built using the algorithm proposed by Parkinson et al. (2008), which those authors demonstrated provided a good match to several statistics of merger trees extracted from the Millennium Simulation. Furthermore, Jiang & van den Bosch (2013) study several algorithms for constructing merger trees and compare their results with trees extracted from N-body simulations. They find that the Parkinson et al. (2008) algorithm is the only algorithm which gives good agreement with the N-body merger trees for all statistics that they consider, and that (given the uncertainties in the construction of N-body merger trees) merger trees constructed using this algorithm are as accurate as those extracted from N-body simulations.

In this Appendix, we explore this algorithm further, comparing it to a set of merger trees extracted from the “MillGas” simulation (kindly provided for this study by the VIRGO Consortium). This simulation has Millennium-like resolution, cosmological parameters $(\Omega_M, \Omega_\Lambda, \Omega_b, H_0, \sigma_8, n_s) = (0.272, 0.728, 0.045, 70.4 \text{ km/s/Mpc}, 0.81, 1.0)$, but utilizes a smaller box of $125/h$ Mpc on a side. Despite the name, the version of MillGas used for this analysis is dark matter-only. Halo finding and merger tree construction is performed using SUBFIND and D-TREES respectively (Jiang et al. 2014).

The primary advantage of this simulation for purposes of our analysis is that a large number (976) of snapshots were stored from the simulation, allowing us to explore how the statistics of merger trees depend on the number of snapshots used in their construction. We find that the merger

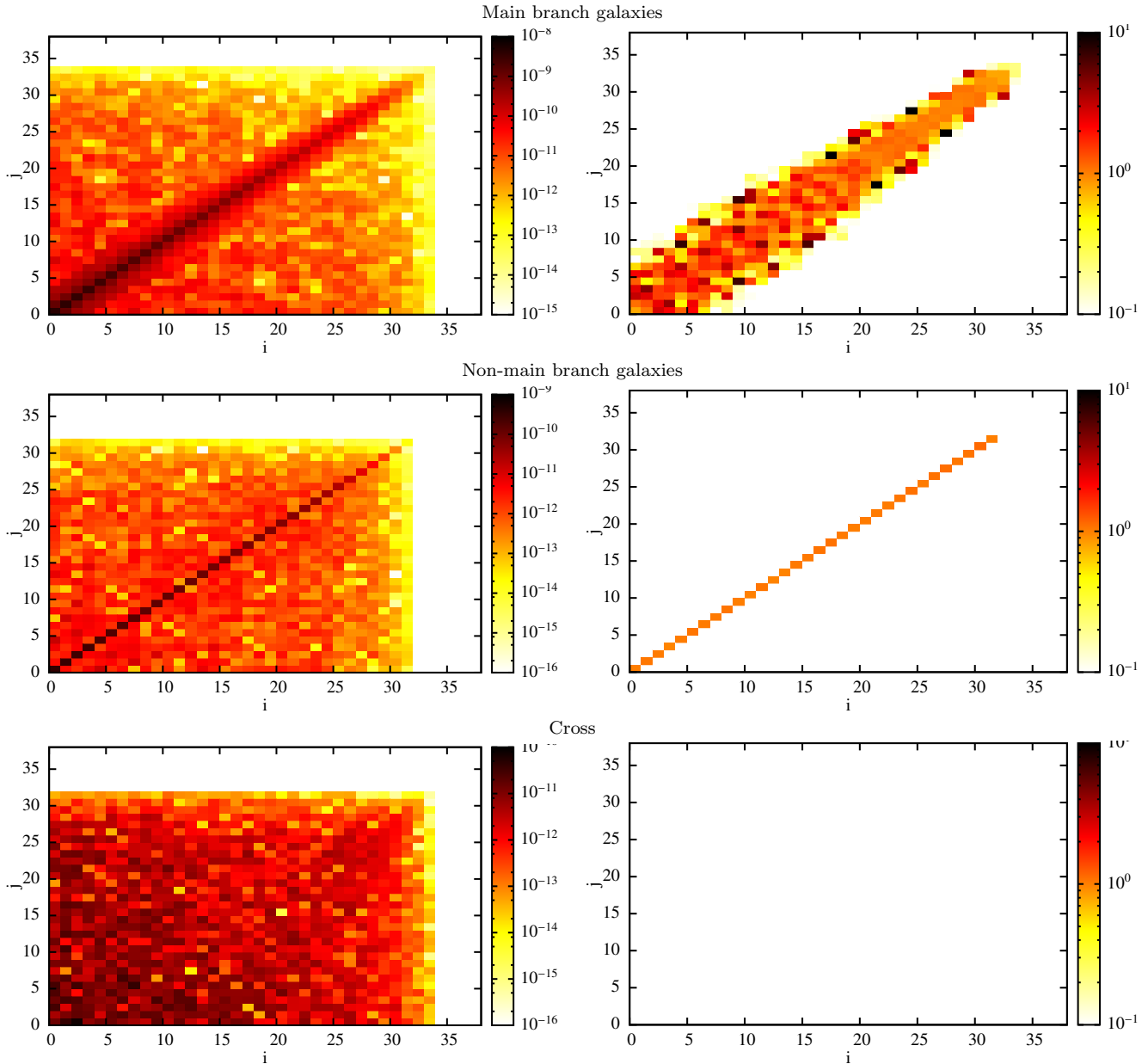


Figure D1. *Left column:* Covariance matrices measured from 1000 realizations of the model stellar mass function are shown for main branch galaxies, non-main branch galaxies, and the covariance between these two populations (i.e. considering the full covariance matrix as a 2×2 block matrix, these are the two diagonal blocks and the off-diagonal block respectively). *Right column:* The ratio of model to measured covariances. White areas indicate zero model covariance.

tree statistics are well-converged with 976 snapshots (as expected; Benson et al. 2012).

Comparing the conditional mass functions (i.e. the distribution of halo masses, M_a , at redshift z_a , conditioned on those halos being progenitors of a halo of mass M_b at the later redshift z_b), we find excellent agreement between N-body and Monte Carlo trees, as did Parkinson et al. (2008). In fact, we find that this agreement extends to the conditional mass functions of the 1st through 4th most massive progenitors (Parkinson et al. (2008) checked up to the 2nd most massive progenitor), and for very small timesteps ($\Delta z = 0.00122$ at $z = 0$ for example).

One issue that we find does make a significant differ-

ence is whether we allow halos in the N-body merger trees to lose mass. As noted by Helly et al. (2003), halos in N-body merger trees can be less massive than their progenitors, while this is not possible by construction in the Parkinson et al. (2008) algorithm. Helly et al. (2003) were forced to confront this issue as their semi-analytic model required monotonically increasing halo masses. They therefore decided to force monotonicity by adding mass to halos which were less massive than the sum of their progenitor masses, making them equal to the sum of the progenitor masses. The N-body merger trees used by Parkinson et al. (2008) were forced to have monotonically increasing halo masses.

While GALACTICUS is entirely capable of handling non-

monotonic halo masses, there is evidence (John Helly, private communication) that some of this reduction in mass is a numerical artifact of the friends-of-friends algorithm which tends to artificially transfer mass from the second most massive progenitor just prior to a merger. Therefore, it is worth exploring the effects of enforcing monotonicity.

Figure E1 shows the conditional mass function of N-body and PCH merger trees for $M_b = 10^{12.22} M_\odot$, $z_b = 0$, $z_a = 0.00122$, 0.0199 , and 0.989 (top, middle, and bottom rows respectively), as a function of mass ratio M_a/M_b . Considering first the left column, in which the N-body trees have not been forced to be monotonically increasing in mass along each branch, the two show very good agreement, even for small time differences, with the following caveats:

- For mass ratios above unity the PCH trees have no progenitors, while N-body trees can;
- There are small offsets in the mass ratio below which the conditional mass function is cut off.

This latter point is not surprising—the cut off arises from the finite mass resolution of the N-body simulations, which we attempt to match by imposing the same mass resolution limit on our Monte Carlo trees. However, there is no reason to expect that the details of the shape of the cut off will be matched—these will depend on precisely how resolution affects tree construction for Monte Carlo trees, and halo finding in the N-body simulations. It is also irrelevant for any careful calculation of galaxy properties which should, of course, check that changes in resolution make no difference to the results.

The right column of Figure E1 shows the same comparison but with monotonically increasing halo masses forced on the N-body merger trees. By construction, the N-body trees now have no progenitors with mass ratios above unity. The agreement with the Monte Carlo trees remains just as good, perhaps slightly better. For example, consider the point for $z_a = 0.0199$ (middle row) in the left column at mass ratio 0.5. The N-body point lies significantly below the Monte Carlo results. However, once monotonically increasing halo masses are enforced these points agree extremely well.

To further explore the effects of enforcing monotonically increasing halo masses we show in Figure E2 three other statistics for the merger trees of halos of mass $10^{13.5} M_\odot$ at $z = 0$. In each case the left column shows the results for N-body trees *without* enforcing monotonically growing halo masses, while the right column *does* enforce this. The top row shows the distribution of formation times (defined as the time at which 50% of the final halo mass was first assembled into a single halo). In this case, agreement between N-body and PCH merger trees is good irrespective of whether halo masses are forced to increase monotonically.

The middle row shows the distribution of last major merger times. This distribution is quite flat, but in the case where N-body trees are not forced to have monotonically increasing halo masses there is a clear difference with PCH trees, such that the N-body trees have last major merger times peaked toward early times. This is to be expected if the friends-of-friends algorithm artificially reduces the masses of halos just prior to merging. If we enforce halo masses to grow monotonically, the distribution for N-body trees now agrees very closely with that from PCH trees.

Finally, we examine the fraction of mass gained by the

halo over its lifetime which can not be accounted for by the masses of merging halos (bottom row). In this case, N-body halos typically gain a large fraction of their mass from this “smooth accretion”, much more so than the corresponding PCH halos, when halo masses are not required to increase monotonically with time. This is again what would be expected if the friends-of-friends algorithm artificially reduced the masses of halos just prior to merging. Enforcing monotonic growth results in distributions of accreted mass that are in excellent agreement between the two sets of merger trees.

In conclusion, once monotonic growth of halo mass is enforced and the N-body trees have sufficiently high time resolution, there is no significant discrepancy between the N-body and PCH trees. This suggests the following:

- (i) the PCH algorithm works extremely well in matching the N-body tree conditional mass functions, including distributions of first through fourth most massive progenitor, formation times, time of last major merger, and fraction of mass accreted sub-resolution;
- (ii) masses of the secondary progenitor are artificially reduced just prior to merging in N-body merger trees.

This latter issue seems to be a result of the friends-of-friends algorithm (John Helly, private communication) and, as we have shown, can be mitigated by forcing the masses of halos in the N-body merger trees to always grow with time along each branch. For this reason, we choose to enforce monotonic growth on halos when computing the model discrepancy in §3.5.3.

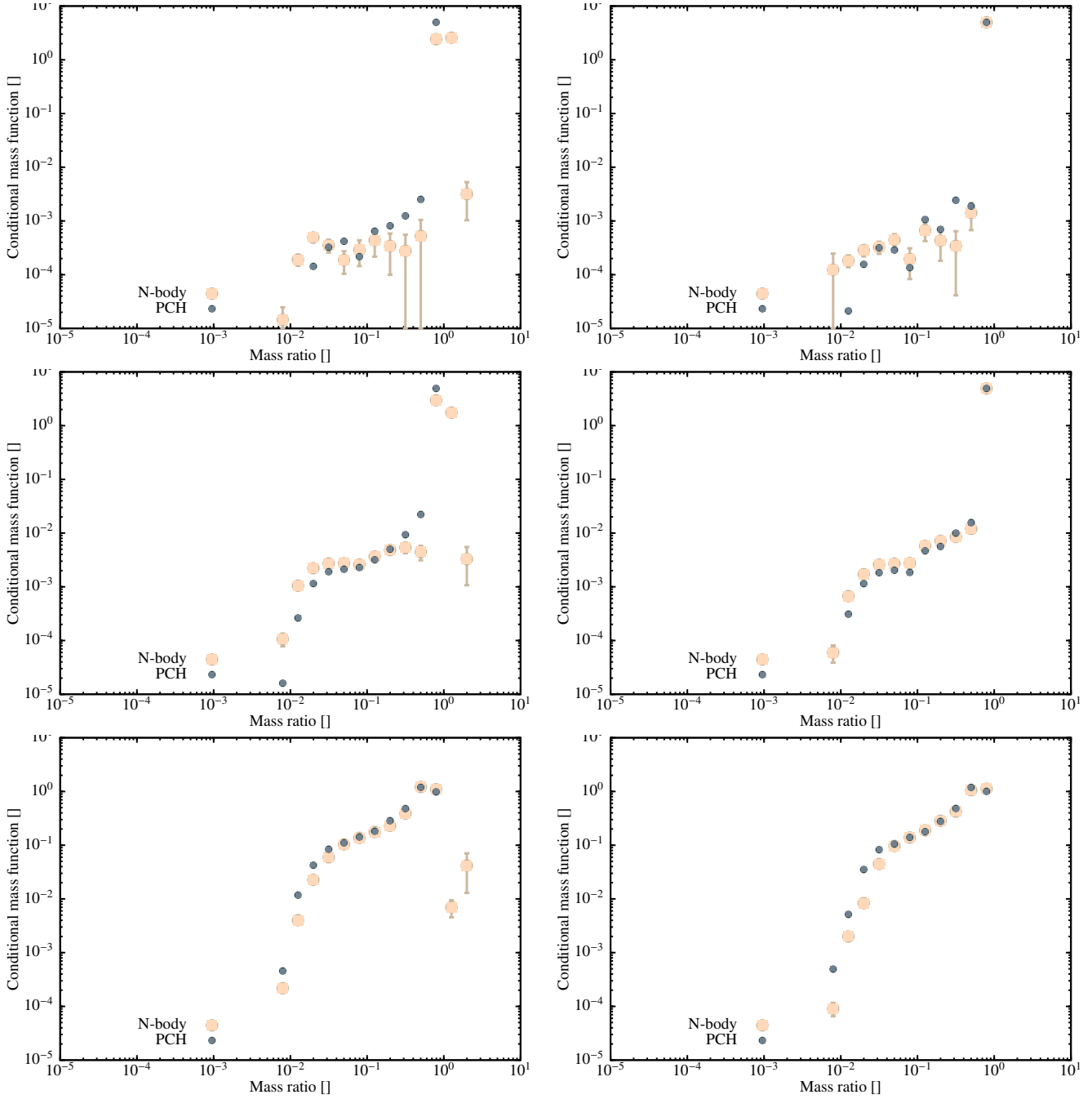


Figure E1. Conditional mass functions from N-body and PCH merger trees for $M_b = 10^{12.22} M_\odot$, $z_b = 0$, $z_a = 0.00122$, 0.0199 , and 0.989 (top, middle, and bottom rows respectively), as a function of mass ratio M_a/M_b . The left column shows the conditional mass function *without* enforcing halo masses to grow monotonically, while the right column shows the conditional mass function *with* monotonicity enforced.

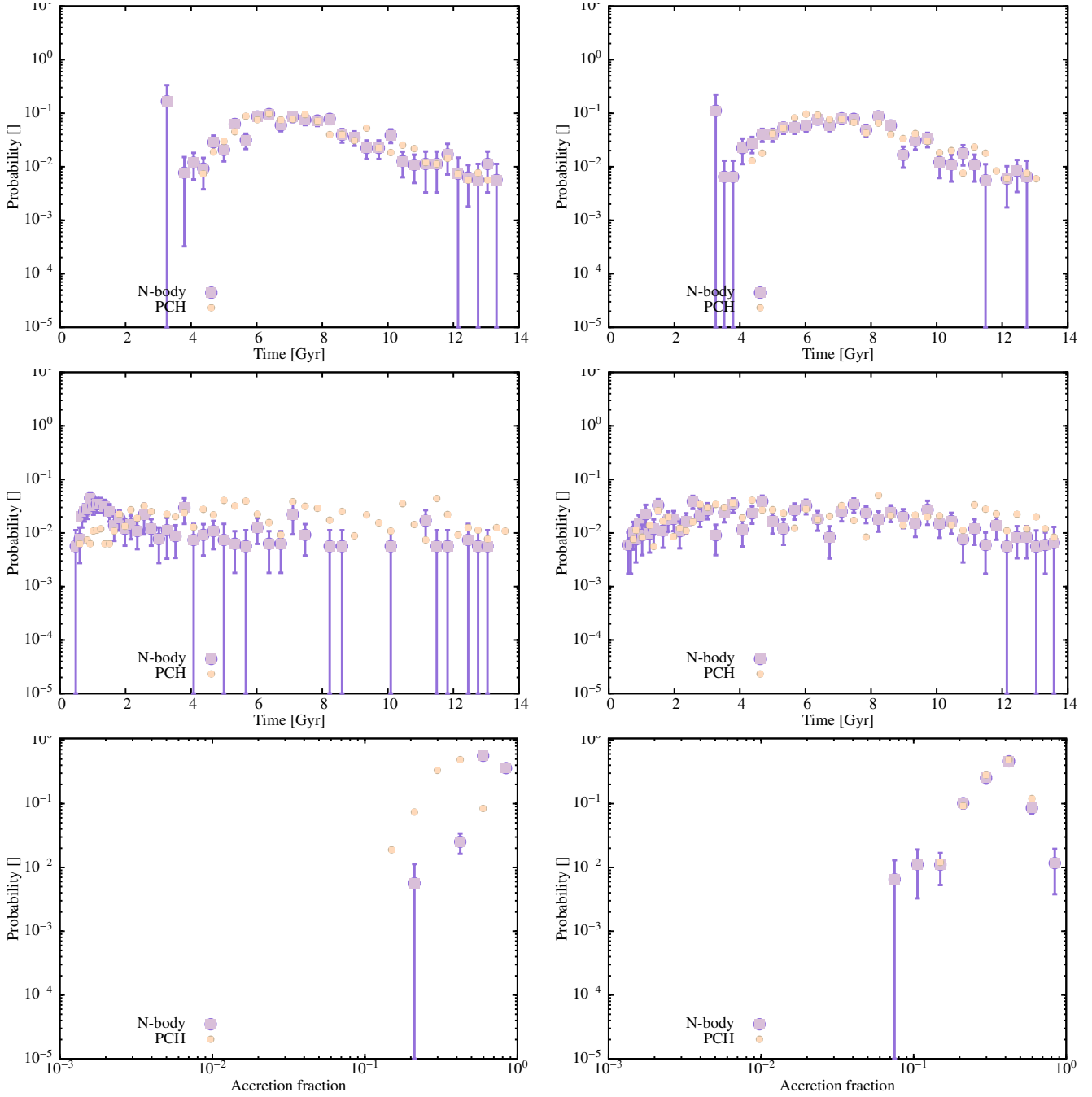


Figure E2. Statistics of the merger trees of halos of mass $10^{13.5} M_{\odot}$ at $z = 0$. The left column shows results for N-body merger trees *without* enforcing monotonically increasing halo masses, while the right column shows results when this *is* enforced. The top row shows the distribution of formation times (defined as the time at which 50% of the final halo mass was assembled into a single halo), the middle row shows the distribution of last major merger times (defined as a merger with mass ratio of 4:1 or lower), and the bottom row shows the distribution of accretion fraction (i.e. the fraction of mass in the final halo which was gained via smooth accretion, and not via mergers with resolved halos).